

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Network Sharing Through Service Outsourcing in Inter-domain IMS Frameworks

Gavole Vitalis Ozianyi

Supervisor:

Neco Ventura



Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Electrical Engineering
UNIVERSITY OF CAPE TOWN

September 2010

Declaration

I declare that this thesis is my own work. Where collaboration with other people has taken place, or material generated by other researchers is included, the parties and/or material are indicated in the acknowledgements or references as appropriate.

This work is being submitted for the Doctor of Philosophy Degree in Electrical Engineering at the University of Cape Town. It has not been submitted to any other university for any other degree or examination.

Gavole Vitalis Ozianyi

Signed by candidate

September 2010

Acknowledgments

I would like to express my sincere gratitude to the following individuals and organisations for their assistance during the course of this project.

Mr. Neco Ventura, for his supervision and guidance throughout the project.

Prof. Rui Aguiar and Susana Sargento of the Institute of Telecommunications at the University of Aveiro for their input and constructive discussions.

Richard Good to whom I am grateful for their advice and constructive feedback.

My fellow colleagues in the Communications Research Group (CRG) at UCT, for the useful discussions during various presentations.

My parents, sisters and brothers, for their love and constant support in every aspect of my life. Thank you.

Synopsis

Mobile operators make capital investments in infrastructure to provide the resources required for service provisioning. The level of investments depends on network dimensioning that aims to balance the supply and demand of resources. Generally the demand for network resources is based on average demand measurements taken over a period of time. The peak values of the average demand estimates are used in dimensioning the resource capacity of the networks. Since demand is not constant, it often happens that the actual demand for resources exceeds the provisioned capacity. When this happens, users are subjected to poor Quality of Experience (QoE) that may manifest in the form of high call blocking rates and low QoS. The likely outcome of this is churning from the network to subscribe with competing operators in anticipation of receiving better service.

Investing in new technologies is considered a credible way of increasing network capacity and efficiency. However, it is only feasible to undertake such a venture after recouping the Capital Expenditure (CAPEX) associated with previous investments. Moreover, network upgrades are not a feasible solution for short-term spikes in the demand for resources.

Resource sharing can be used as a short-term solution to the imbalance between the supply and demand of network resources. Resource sharing enables operators to provide services to their subscribers using networks belonging to other operators. Resource sharing in mobile networks is increasingly becoming an option for operators to provide service to their subscribers. In this thesis we explore a mechanism for sharing access network resources that utilises negotiable short-term Service Level Agreements (SLA) that can easily adapt to changing network conditions. Through this mechanism operators of resource constrained networks may use near real-time dynamic SLAs to negotiate network access services for their subscribers. We refer to this form of resource sharing

as 'Service Outsourcing'.

Service outsourcing is characterised by a wireless overlap between the access networks of the involved operators. Without outsourcing, during peak periods the network would block incoming service requests when the bandwidth available on the home access network is insufficient. Moreover, the network may drop or degrade the performance levels of active sessions. However, through service outsourcing the home network may request resources from a visited network that shares a wireless overlap to accommodate the requests. For example the home operator of a 3G mobile network may request resources from an LTE mobile network. Service outsourcing would take place if resources are granted by the LTE network to admit the requested sessions. The affected calls and sessions would be transferred to the visited network using handover schemes like mobile IP and SIP based session continuity. The latter scheme is used in this thesis.

Service outsourcing SLAs stipulate the sharing of revenue accruing from user payments between operators. The main characteristic of such SLAs is the need to maintain a predictable service charge for users of outsourced sessions. At the end of service delivery the user of an outsourced session would be charged the same amount like a user of similar services that were provisioned using the home access network. Service outsourcing SLAs are session based and negotiated at session setup time.

Service outsourcing agreements involve the following parameters: an outsource duration, a specific QoS level and an offer price. The outsource duration sets a lower bound on the time an active outsourced session would be served on the visited access network. This duration prevents operators from handing a session back to the home network in an attempt to avoid the implications of outsourcing. The QoS level is guided by the bandwidth requirements of the outsourced session: the visited operator's access network should be able to allocate the required bandwidth for outsourcing to occur. The offer price determines income share proportions for each operator. Each operator prefers an offer price value that enables them to maximise on outsourcing benefits. Generally operators can maximise on revenue shares from service outsourcing by applying strategies aimed at manipulating values of the specified outsourcing parameters to their favour.

In an environment with multiple network operators, those with abundant network capacity may compete to provide services to outsourced sessions. Multi-operator scenarios require bidding mechanisms to facilitate speedy negotiations of service outsourcing SLAs. Moreover, technical procedures e.g., charging platforms, and interfaces for interaction

between operators are required.

The 3GPP standardised the IP Multimedia Subsystem (IMS), in the release 5 of its specifications, to provide convergence between mobile networks and the Internet. Since then there have been new releases of specifications for functions related to IMS like policy and charging control. Currently release 9 of the specifications provides guidelines for charging and policy control. Network equipment vendors and operators follow these specifications to achieve inter-operability of their systems. The research community also uses these standards to guarantee activity that is up to date with industry practise. However, it should be noted that the development of individual components of the IMS is vendor specific. This thesis utilises development work that is part of the open IMS research, which provides a framework for the development and testing of telecommunication mechanisms and services.

Open Source IMS software that provides the SIP Call Session Control Functions (CSCF) and a Home Subscriber Server (HSS) is available for research and development purposes. Service outsourcing in IMS benefits from the existence of this framework. However, we note that the Open IMS framework does not include a charging and billing platform. Service outsourcing requires a standards compliant comprehensive charging system for the IMS. It is for this reason that we develop a charging framework for the IMS. The charging framework is designed in line with 3GPP standards and to inter-operate with the Open IMS testbed from Fraunhofer Fokus. Moreover, we design and develop a service outsourcing framework for the IMS.

This thesis investigates service outsourcing with the aim of establishing through analytical methods and testbed experiments how real-time session based short-term SLAs may be used to facilitate resource sharing. At the end of the thesis, we present results obtained from testbed experiments to show the effect of service outsourcing on the incomes of involved operators. The results will show the impact of varying the offer price and outsource durations on outsourcing incomes. The results also show the effect of the outsource duration and average call durations on the number and pattern of sessions that are blocked or outsourced by the home network when it lacks bandwidth to admit arriving service requests. The results highlight the benefits and implications of participating in outsourcing by the home and visited operators. Moreover, the testbed results are used as proof of concept of the applicability of service outsourcing in practical environments.

Contents

Declaration	i
Acknowledgments	ii
Synopsis	iii
List of Figures	xiii
List of Tables	xvii
List of Acronyms	xviii
1 Introduction	1
1.1 Service Level Agreements	5
1.2 Network Sharing Solutions	6
1.3 Service Outsourcing	7
1.4 IP Multimedia Subsystem (IMS)	11
1.4.1 Service Outsourcing in IMS	12
1.4.2 Charging in IMS	13
1.4.3 Service Outsourcing Signalling and Scalability	15
1.5 Service Outsourcing Benefits and Challenges	16
1.6 Thesis Objectives	17

1.7	Thesis Scope and Limitations	18
1.8	Thesis Contributions	19
1.9	Thesis Outline	21
2	Literature Review	23
2.1	Introduction	23
2.2	Next Generation Wireless Access Networks	25
2.3	Network Sharing Solutions	29
2.3.1	Site Sharing	29
2.3.2	Radio Access Network (RAN) Sharing	29
2.3.3	Full Network Sharing	30
2.4	Service Level Agreements (SLA)	34
2.5	Internet Economics	36
2.5.1	Pricing and Charging Schemes	36
2.5.2	Pricing Profiles	38
2.5.3	Demand and Supply of Network Resources	39
2.6	Service Provision Architectures	41
2.6.1	Service Control	42
2.6.2	Service Phases	42
2.7	Discussion	44
3	Background of Charging Control in IMS	46
3.1	Charging and Billing	48
3.1.1	Diameter Protocol	49
3.1.2	Diameter Credit Control Application	50
3.1.3	Flow Level (volume) Charging	50
3.1.4	Content Charging	51

3.2	IP Multimedia Subsystem	51
3.2.1	Proxy Call Session Control Function (P-CSCF)	52
3.2.2	Serving Call Session Control Function (S-CSCF)	52
3.2.3	Interrogating Call Session Control Function (I-CSCF)	53
3.2.4	Home Subscriber Server (HSS)	53
3.2.5	Home and Visited Domains	53
3.2.6	IMS Charging Control	54
3.3	Evolved Packet System	56
3.3.1	Roaming in EPS	58
3.4	Conclusions	58
4	Network Capacity Expansion Through Service Outsourcing	60
4.1	Service Outsourcing Concepts	60
4.1.1	Service Outsourcing Scenario	61
4.1.2	Requirements and Conditions	61
4.1.3	Service Outsourcing SLA	62
4.1.4	Service Outsourcing Transactions	63
4.2	Outsourcing Architecture	67
4.2.1	User Agent	68
4.2.2	Network Agent	68
4.2.3	Multiple Candidate Networks	69
4.2.4	Combined Outsourcing Transactions	69
4.3	Pricing of Network Resources	70
4.3.1	Custodian Network Outsource Profit	71
4.3.2	Candidate Network Outsource Profit	72
4.3.3	Financial Implications of Service Outsourcing	72
4.4	Network Resource Supply and Demand Model	77
4.5	Conclusions	78

5	Service Outsourcing in IMS Frameworks	79
5.1	IMS Service Outsourcing Architecture	79
5.1.1	Outsourcing Network Agent	80
5.1.2	Outsourcing User Agent	82
5.1.3	Outsourcing Setup Phase	83
5.1.4	Outsourcing Execution Phase	84
5.2	Charging in IMS	86
5.2.1	Offline and Online Charging Systems	87
5.2.2	VoIP Charging System	91
5.2.3	IPTV Charging System	93
5.2.4	Volume Charging	95
5.2.5	Charging for Service Outsourcing	96
5.3	Summary	97
6	Outsourcing Evaluation Framework	98
6.1	Open Source IMS	99
6.2	UCT IMS Client	100
6.3	UCT IPTV Framework	101
6.3.1	Media Server	101
6.3.2	HSS Configuration	102
6.4	Outsourcing Framework	103
6.4.1	Outsourcing User Agent	105
6.4.2	Outsourcing network agent	106
6.4.3	Charging Framework	107
6.5	Summary	113

7	Testing and validation	114
7.1	Proof of Concept Tests	114
7.1.1	IMS Charging System	115
7.1.2	Service Outsourcing: Custodian Network - Setup Phase	116
7.1.3	Service Outsourcing: Candidate Network - Execution Phase	119
7.1.4	Outsourcing Network Agent Messages	121
7.2	Session Establishment Delay	124
7.2.1	Service Level Charging	124
7.2.2	Policy and Charging Control on the Media Plane	124
7.2.3	Service Outsourcing - Sessions Admitted on Custodian Network	127
7.2.4	Service Outsourcing - Sessions Outsourced to the Candidate Network	128
7.2.5	Service Outsourcing - A General Scenario	129
7.3	Distribution of Custodian, Candidate and Blocked Calls	132
7.3.1	Treatment at Call Arrival - Admitted, Outsourced and Rejected Sessions	134
7.4	Network Capacity Utilisation - Custodian and Candidate Network Sessions	141
7.4.1	Scenario 1a	141
7.4.2	Scenario 1b	141
7.4.3	Scenario 2a	142
7.4.4	Scenario 2b	144
7.4.5	Scenario 3a	145
7.4.6	Scenario 3b	146
7.4.7	Scenario 4	146
7.5	Virtual Capacity Utilisation	147
7.6	Outsourcing Benefits	149
7.7	Summary	153

8	Conclusions and Recommendations	155
8.1	Conclusions	155
8.1.1	Network Churn	155
8.1.2	Network Sharing Solutions	156
8.1.3	Service Outsourcing	156
8.1.4	Operator Attitudes	157
8.1.5	Service Outsourcing in IP multimedia Subsystem	157
8.1.6	Charging in IMS Frameworks	157
8.1.7	Testbed Implementation	157
8.1.8	Evaluation Tests	158
8.2	Future work	159
8.2.1	Impact of Outsourcing on the Candidate Network	159
8.2.2	Practical Deployment	160
8.2.3	Application to other Environments	160
8.2.4	Security and Privacy	160
	Bibliography	161
A	Wireless Next Generation Networks	i
A.1	Heterogeneous Next Generation Networks	i
A.1.1	WLAN Access	i
A.1.2	WiMax Access	ii
A.1.3	The 3GPP Evolved UTRAN (E-UTRAN)	iii
A.2	Differentiated Services Architecture	v
B	Charging control in IP Multimedia Subsystem	vii
B.1	IMS session control	vii
B.2	IMS policy and charging control	ix

B.2.1	Online charging	ix
B.2.2	Offline charging	ix
B.2.3	IMS Session Based Charging	x
B.2.4	Policy and charging control over Gx reference point	x
B.2.5	Policy and charging control over Rx reference point	xii
B.2.6	Policy and charging control over S9 reference point	xii
B.2.7	Real-time Transport Protocol (RTP)	xiii
C	Service outsourcing transactions	xiv
C.1	Numerical data	xiv
D	Outsourcing procedures in IMS	xv
D.1	IMS Registration	xv
D.2	Session establishment	xvi
E	Evaluation Framework	xvii
E.1	Source Code	xvii
E.1.1	UCT IMS Client	xvii
E.1.2	IPTV Application Server	xvii
E.1.3	RTSP Media Server	xix
E.1.4	Outsourcing User Agent	xx
E.1.5	Network Agent	xx
E.1.6	Charging Framework	xx
F	Accompanying CD-ROM	xxii

List of Figures

1.1	Service outsourcing involving three networks	8
2.1	A simplified version of the evolved packet system	27
2.2	Site based sharing solution	29
2.3	Radio Access Network Sharing	30
2.4	Full network sharing	30
3.1	Policy and charging control architecture	47
3.2	Basic entities of the IMS architecture	52
3.3	Service level charging in IMS	54
3.4	TISPAN IMS IPTV architecture	56
3.5	Creation of PCC rules in EPS	57
3.6	S9 policy and charging control reference architecture	58
4.1	Service outsourcing general architecture	61
4.2	Service outsourcing - setup phase	64
4.3	Transactions during the execution phase	65
4.4	Service outsourcing termination phase	66
4.5	Outsourcing network agent	68
4.6	Income variation due to outsourcing	74
4.7	Sliding outsource duration (T) with congestion thresholds	75

5.1	IMS service outsourcing architecture	80
5.2	IMS outsourcing architecture	81
5.3	Architectural design of the network agent	82
5.4	Outsourcing user agent	83
5.5	IMS session outsource negotiation	85
5.6	Charging system layout	86
5.7	Functional architecture of the charging AS	87
5.8	Structure of the offline charging system	89
5.9	Functional architecture of the online charging system	90
5.10	Online credit control	92
5.11	VoIP online charging	93
5.12	Charging for IPTV services	94
5.13	Architectural layout of the IPTV AS	95
5.14	Outsource charging design layout	96
6.1	Layout of the Open IMS core	99
6.2	UCT IMS client	100
6.3	Functional structure of the IPTV application server	101
6.4	Provisioning the candidate network in IMS	102
6.5	Authorisation of service outsourcing for IMS users	103
6.6	Definition of IPTV application server at the HSS	104
6.7	Definition of service point triggers for IPTV	105
6.8	Layout of the evaluation framework	106
6.9	Charging Framework	109
6.10	Prepaid and post paid selections for online and offline charging	110
6.11	Service rating information	111
6.12	Subscriber account balance information	111

7.1	Credit control by the online charging system	115
7.2	IPTV volume charging	116
7.3	IPTV Session setup delay	125
7.4	Session setup delay with PCC procedures	126
7.5	Setup delays for sessions admitted on the custodian network	127
7.6	Session setup delay with and without outsourcing	128
7.7	Setup delays for outsourced sessions	129
7.8	Delays for outsourced and admitted sessions	130
7.9	Call setup delays for a comprehensive case - scenario 1a	131
7.10	Call setup delays for a comprehensive case - scenario 2a	131
7.11	Call arrival: Admitted, outsourced and blocked sessions - scenario 1a . . .	134
7.12	Call arrival: admitted, outsourced and blocked sessions - scenario 1b . . .	135
7.13	Call arrival: admitted, outsourced and blocked sessions - scenario 2a . . .	136
7.14	Call arrival: admitted, outsourced and blocked sessions - scenario 2b . . .	137
7.15	Call arrival: admitted, outsourced and blocked sessions - scenario 3a . . .	138
7.16	Call arrival: admitted, outsourced and blocked sessions - scenario 3b . . .	139
7.17	Call arrival: admitted, outsourced and blocked sessions - scenario 4	140
7.18	The number of custodian and candidate network calls - scenario 1a	142
7.19	The number of custodian and candidate network calls - scenario 1b	143
7.20	The number of custodian and candidate network calls - scenario 2a	143
7.21	The number of custodian and candidate network calls - scenario 2b	144
7.22	The number of custodian and candidate network calls - scenario 3a	145
7.23	The number of custodian and candidate network calls - scenario 3b	146
7.24	The number of custodian and candidate network calls - scenario 4	147
7.25	Virtual capacity utilisation by the custodian network- scenario 1a	148
7.26	Virtual capacity utilisation by the custodian network- scenario 3a	149

7.27	Virtual capacity utilisation by the custodian network- scenario 4	150
7.28	Expected income at call arrival - scenario 1a	150
7.29	Expected income at call arrival - scenario 1b	151
7.30	Expected income at call arrival - scenario 2a	151
7.31	Expected income at call arrival - scenario 2b	152
7.32	Expected income at call arrival - scenario 3a	153
7.33	Expected income at call arrival - scenario 3b	154
7.34	Expected income at call arrival - scenario 4	154
A.1	WiMax network reference Model	iv
B.1	IMS session control	viii
B.2	Service level charging in IMS	x
B.3	Policy and charging control architecture	xi
B.4	S9 policy and charging control reference architecture	xiii
D.1	Registration in the IMS	xv
D.2	IMS sessions establishment	xvi
E.1	IPTV iFC configuration at the HSS	xix
E.2	The HSS service profile configuration	xix

List of Tables

2.1	The evolution of GSM wireless networks	26
2.3	Wireless local area networks	28
C.1	Numerical data	xiv

University of Cape Town

List of Acronyms

3G	Third Generation
3GPP	3rd Generation Partnership Project
AAA	Authentication Authorisation and Accounting
ACA	Accounting Answer
ACR	Accounting Request
AF	Application Function
AP	Access Point
A-GW	Access Gateway
AVP	Attribute Value Pair
BBERF	Bearer-Binding and Event-Reporting Function
CAPEX	Capital Expenditure
CCA	Credit Control Answer
CCR	Credit Control Request
CDF	Charging Data Function
CGF	Charging Gateway Function
CSCF	Call Session Control Function
CTF	Charging Trigger Function
DiffServ	Differentiated Services
DNS	Domain Naming System
ePDG	Evolved PDG
EPS	Evolved Packet System
E-UTRAN	Evolved UTRAN
GGSN	GPRS Gateway Support Node
GPRS	General Packet Radio Service

GSM	Global System for Mobile communications
GW	Gateway
FTP	File Transfer Protocol
HLR	Home Location Register
HPCRF	Home Policy and Charging Rules Function
HPLMN	Home Public Land Mobile Network
HSS	Home Subscriber Server
I-CSCF	Interrogating Call Session Control Function
IETF	Internet Engineering Task Force
IntServ	Integrated Services
IP-CAN	IP Connectivity Access Network
IPTV	IP Television
ISC	IMS Service Control
ISP	Internet Service Provider
LAN	Local Area Network
LTE	Long Term Evolution
MAC	Media Access Control
MME	Mobility Management Entity
MMS	Multimedia Messaging Service
MNO	Mobile Network Operator
MVNO	Mobile Virtual Network Operators
NGN	Next Generation Network
OFCS	Offline Charging System
OCF	Online Charging Function
OCM	Outsource Cost Margin
OCS	Online Charging System
OPEX	Operating Expenditure
OSA	Open Services Architecture
PCC	Policy and Charging Control
PCEF	Policy and Charging Enforcement Function
P-CSCF	Proxy Call Session Control Function
PCRF	Policy and Charging Rules Function
PDG	Packet Data Gateway
PDN	Packet Data Network

PDN GW	PDN Gateway
QoE	Quality of Experience
QoS	Quality of Service
RADIUS	Remote Dial-In User Service
RF	Rating Function
RTP	Real Time Protocol
S-CSCF	Serving Call Session Control Function
SDP	Session Description Protocol
S-GW	Serving Gateway
SIP	Session Initiation Protocol
SLA	Service Level Agreement
SMTP	Simple Mail Transfer Protocol
SPT	Service Point Trigger
TISPAN	Telecommunications and Internet converged Services and Protocols for Advanced Networking
ToD	Time of Day
UA	User Agent
UE	User Equipment
URI	Universal Resource Identifier
UMTS	Universal Mobile Telephone System
UTRAN	UMTS Terrestrial Radio Access Network
VHE	Virtual Home Environment
VPCRF	Visited PCRF
VPLMN	Visited Public Land Mobile Network
VoD	Video on Demand
WiMax	World wide interoperability for Microwave access
WLAN	Wireless LAN

Chapter 1

Introduction

Mobile and wireless communications rely on radio frequency spectrum resources to provide the bandwidth required for both signalling and media traffic. As the amount of bandwidth available on such networks is finite only a limited number of users may be served simultaneously. The cost of the frequency spectrum licenses, capital expenditure related to site infrastructure and operational costs contribute to the overall cost of providing mobile communication services. Networks that operate at close to full capacity are likely to generate more income, while underutilised networks may incur losses when the cost of resource provisioning exceeds returns from service delivery. When an available network's capacity is constrained network efficiency is compromised, which culminates in poor network operation. During periods of congestion, new service requests as well as handover requests, due to user mobility, are rejected.

Unchecked levels of service denial may translate to long-term user dissatisfaction with operator services.

User experience with network services directly influences their decision to maintain subscription with an operator. Dissatisfied users may opt to terminate subscriptions (resulting in network churn) and seek an alternative, often competing, operator. High network churn rates will affect operator revenue flows; moreover, signing up new subscribers is an expensive exercise. User satisfaction may be tied to, among other factors, the number of unsuccessful call establishment attempts. Thus availability of sufficient resources is an essential condition to be met by the network in order to minimise call blocking rates or QoS degradation and the associated churn rate [1]. The user Quality of Experience

(QoE) with telecommunication services depends on the extent to which QoS requirements of the services are met. Sufficient network resources are needed to meet service QoS requirements and sustain the required QoE. QoE relates to user perception of the level of services rendered to them.

Network planning takes care of the supply and demand of resources. The supply of resources typically aims at meeting the peak value of demand averaged over a certain time frame. When the demand exceeds the available resources, service degradation occurs as a result of network congestion. This is depicted by unbounded delay levels, high call blocking rates or packet loss. Telecommunication network operators deploy infrastructure with the aim of offering a limited set of services to their subscribers. However, new appealing services are created continuously, thus making it necessary for these networks to support services for which they are not optimised. Such services often consume greater bandwidth; hence more resources are needed to achieve the required QoE levels. The efficiency and effectiveness of network operation would decline if these services were deployed on legacy networks. Network operators would upgrade their infrastructure to provision more resources and gain from revenue potentials that arise from offering these new services.

Upgrading of networks to provision more resources is a suitable long term solution. However it may not solve challenges associated with short term spikes in the demand for resources. Ideally, operators would prefer to operate networks close to full capacity. Abundant resources are required to accommodate peak service requests. Thus it is important, through network planning, to provision sufficient network resources to meet average demands and operate efficiently. During periods of high congestion the network may experience continuous arrival of new service requests. As mentioned earlier, blocking of new requests or degradation of active flows to accommodate new flows results in user dissatisfaction. Over provisioning of network resources may not be feasible in terms of revenue targets required to offset the cost of resource provisioning. Thus, mechanisms for solving the resource demand vs. supply and the associated cost paradox are needed. Network resource sharing among competing operators may be an appropriate solution to meeting short term peak demands for resources.

Network resource sharing scenarios differ according to adopted business models. Network roaming is a form of resource sharing that enables subscribers to access services by connecting to visited networks belonging to operators with a roaming agreement with

their home operator. In conventional roaming the subscriber is often faced with higher service usage fees associated with meeting the full service cost on both home and visited networks.

The following questions arise: Is it possible to facilitate resource sharing amongst competing network operators to provide services to users at no additional cost? Can such a business model be beneficial to all network operators involved? Are there any short-term and long-term financial implications for involved operators?

To facilitate ubiquitous computing in the mobile domain new business models for resource sharing are needed. This thesis proposes service outsourcing aimed at benefiting competing network operators through resource sharing. In service outsourcing, call blocking or rejection of user service requests by an overloaded network may be minimised by utilising another operator's network to transport user traffic. For the success of such an ubiquitous system, higher service usage charges should not be incurred for using visited operator networks.

The motivation for service outsourcing lies in the following benefits:

- overall enhanced user service experience with regard to minimal call blocking
- facilitation of a best charged and connected ubiquitous computing experience
- minimisation of network churn
- overall improved income for involved operators
- increased operator savings with regard to deployed network infrastructure

Service outsourcing enables a network experiencing resource constraints on a given access network to handover incoming service requests to an overlapping access network belonging to a competing operator. When a service request is received, instead of blocking the request an outsourcing negotiation would be initiated with a federated network to offer services to the user. The requested service would be delivered via the federated network at no additional cost to the user. This condition dictates sharing of revenue accruing from user payments between involved operators.

In this thesis we consider the fact that minimisation of network churn directly relates to enhanced user satisfaction. This applies especially when user satisfaction is influenced by

factors such as the number of unsuccessful service access or call establishment attempts. Although service outsourcing dictates sharing of user payments between involved operators, the overall income for operators could increase with the advantage that both networks would not increase resources (which is always expensive and unfeasible in the short to medium term) but simply shared momentary peak requests.

A question that arises is, under what circumstances should two (or more) operators engage in such co-operation? The answer revolves around the prices that operators agree to charge for a service outsource. If the price is too high, thus resulting in an operator incurring high costs, it would be pointless to engage in such an agreement. The proportion of revenue accruing from user payments to be apportioned to every involved operator depends on the agreements. Resource sharing transactions aimed at meeting sudden overshoots in the demand for resources can be supported with short term income apportioning agreements.

In the long term, revenue sharing may have an impact on the net incomes of operators. Through an analysis of the financial features related to service outsourcing the home operator may decide to upgrade the network, after reaching a threshold point, to boost the capacity and avail more resources to home subscribers. On the other hand, the visited operator may attempt to sign up new subscribers in regions with sustained outsourcing activity.

Service outsourcing agreements require critical parameters to establish the proportion of shared revenue for every outsourced session. The agreements would also define the duration of an active outsource transaction, and conditions under which an outsourced session should be passed back to the home network. Unlike the case of conventional wireless roaming, service outsourcing requires radio reachability of access networks of the home and visited operators by the user equipment. A wireless overlap of the two networks is therefore a pre-requisite. Moreover, the potential visited network should have sufficient resources to accommodate traffic from the outsourced session. Roaming is a resource sharing mechanism primarily aimed at solving lack of network coverage problems, whereas the primary goal of service outsourcing is to increase network capacity.

1.1 Service Level Agreements

Commercial telecommunications are guided by agreements between service providers and their customers. These service level agreements stipulate the characteristics of the service to be offered to subscribers in terms of service availability, quality and cost. Service availability considers geographical location of the served area, periods of service availability etc. Service quality would likely relate to a baseline market standard or previous user experience. Service cost would be influenced by many factors including regulatory controls, market competition, service provision costs including CAPEX and OPEX, and operator revenue targets.

User satisfaction with operator services has a strong relation to SLAs. In a market with many operators and service providers, operators would strive to meet SLA conditions to minimise the chances of losing subscribers to their competitors. It is however common for operators and service providers to create long term binding contracts that make it hard or impossible for dissatisfied users to churn. This prevents such users from switching to other operators in the short term; however, they may still seek services from different operators at the expiry of contracts. They also may seek value added services from other operators. Common causes of subscriber dissatisfaction include call blocking due to network congestion and lack of network coverage in some locations.

Network Churn

The act of dissatisfied subscribers leaving an operator's network to seek services from other operators or providers is known as network churn. Network churn works against provider efforts of signing up new subscribers. Subscriber acquisition involves huge marketing effort and is an expensive exercise. Uncontrolled levels of network churn will reduce subscriber numbers, which are needed for various reasons. A network with many subscribers would be able to distribute costs related to capital investment and OPEX among many users; thus making it possible to lower service delivery costs. Moreover, such a network would easily earn more income and be profitable and competitive.

To minimise network churn levels networks require abundant resources to prevent call blocking and offer coverage in areas where their subscribers may be located. High capital investments are required to provide wide network coverage and to meet expected peak

service demand. In very competitive markets profit margins are likely to be low, thus operators need to reduce capital and operational costs in order to remain competitive during the lifetime of a given technology. Network sharing may provide operators with options for cost minimisation. Increased competition leads to shrinking profit margins for retail providers; thus cost cutting is a major differentiating factor.

1.2 Network Sharing Solutions

Financial commitments by network operators in terms of licence fees, equipment costs, network installation and maintenance costs have forced some operators to resort to network sharing as a means of reducing CAPEX and OPEX when rolling out new services [2]. Stiff competition in addition to high capital expenditure may result in reduced profit margins; through network sharing, operators can reduce costs related to infrastructure provisioning and maintenance. Network sharing may be considered a potential competition inhibitor amongst operators. However, to foster competitive markets the network infrastructure may be provisioned by independent “neutral hosts” who own infrastructure and provide coverage on behalf of operators. This closely follows the wholesale retail model incumbent operators the world over have been forced to adopt.

There are different options available to network operators for network sharing. Network infrastructure sharing may range from simple sharing of cellular towers to full sharing of core equipment. For example, Mobile Virtual Network Operators (MVNOs) own little infrastructure other than a Home Location Register (HLR) and billing system; thus they depend on other operators for network infrastructure. The following sharing solutions have been defined by operators: site and passive RAN sharing; shared RAN; shared core network; full network sharing; geographical split [2]. The important aspect is the existence of a technical and financial architecture that permits co-operation among operators in network sharing. Moreover, competition among independent service providers should remain uninhibited.

Network sharing may be used to increase operator coverage. Additionally or alternatively it may be used to improve resource capacity. The levels of CAPEX savings resulting from network sharing depend on the motivation for sharing. Greater levels of CAPEX saving are achievable in the case of network coverage improvement as compared to capacity improvement. Networks experience periods of low and peak demand at different times

of the day. Generally network capital requirements are a function of the peak demand, while revenue is a function of the average demand [2]. Hence, to achieve a balance between investments and service demand it would be better to dimension networks to meet average resource demands and use network sharing to meet peak demands. The benefits of network sharing are more evident when considering product life cycles. New network technologies will experience growth in different phases and eventually mature. A decline phase will follow, after which enhancements to existing networks and services will be required to sustain required income levels.

Network sharing requires a good regulatory framework to highlight the separation of network and service layers. Thus, network operators can focus on innovation and cost reduction, while service providers dwell on creating new business models and service concepts. There should be numerous specialised service providers, but consolidated, efficient and stable underlying networks. Stringent regulatory structures will prevent the rise of monopolies and occurrence of anti-competitive behaviour. Technical limitations and existing business environments may make it very expensive for multiple operators to run independent networks. This arises in cases where physical resources are under regulator control or a monopolistic provider, e.g., a highway tunnel, airport etc. Installation of network equipment may be disruptive and subject to high rental and other fees, thus making it expensive for network providers. In rural areas with low user densities (implying inadequate demand levels) the capital costs of independent networks may prohibit mass service roll out. Through network sharing, network operators may minimise CAPEX and be able to meet average as well as peak demands for services.

1.3 Service Outsourcing

Service outsourcing would enable operators to share network resources by establishing short time frame SLAs. The partitioning and allocation of network resources would occur in real-time. Service outsourcing would occur between competing network operators to allow a home (custodian) network experiencing resource constraints to transfer service requests to a visited (candidate) network. During normal network operation when new requests for resources are received at the home access network the operator would grant requested resources if there was sufficient capacity. Admitting a new call to a congested network may compromise QoS guarantees of existing calls. This will impact negatively

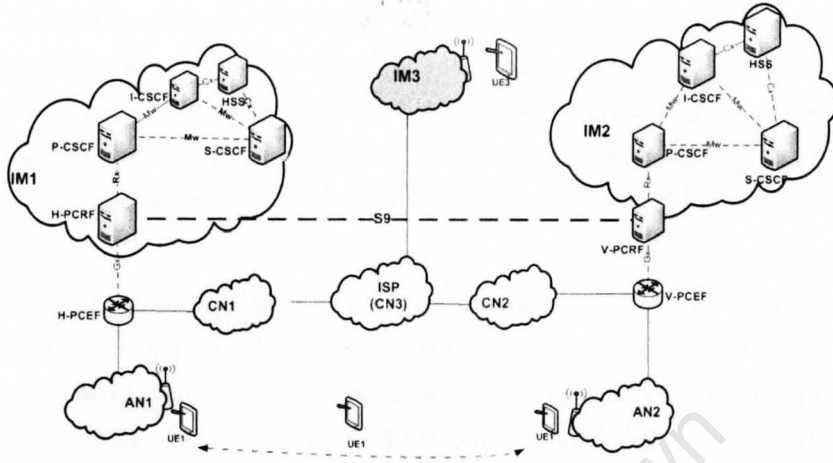


Figure 1.1: Service outsourcing involving three networks

on the user QoE. Rejecting of new resource requests on the congested network may be the appropriate option, but as mentioned earlier it will impact on the user QoE if satisfaction is tied to the number of failed call attempts. Service outsourcing would enable the custodian operator to negotiate network transport services for users of such services on candidate operator networks. Figure 1.1 illustrates service outsourcing involving two network operators.

In the figure a service request is sent from the User Equipment (UE1) towards the home network IM1. The request cannot be admitted due to insufficient capacity. Instead of completely blocking the request an outsourcing requests is initiated by IM1 and sent to a reachable network IM2. After negotiations to establish an outsourcing agreement UE1 is handed over to IM2, where network connectivity and transport services are provided on the new access network (AN2).

Service outsourcing shares a few characteristics with site sharing, RAN sharing and CN sharing solutions. Resource capacity improvement is the main reason for service outsourcing. However, several conditions and requirements should be met for service outsourcing to be successful. For example resource pricing strategies are required to achieve meaningful resource sharing. Technology requirements for service outsourcing include the existence of a wireless overlap involving custodian and candidate operator networks at the user's geographical location. Individual operators may have independent networks with unshared equipment and dedicated radio frequencies. After performing negotiations to establish a service outsourcing SLA, the operators will perform establish-

ment of an outsourcing session. This will involve a handover of the user to the selected visited network. Upon successful establishment of an outsourcing session, call or session control will be handled by both home and visited operators, but communication with the mobile equipment will occur over the visited operator's network. Logically, the custodian and visited networks would share network resources and equipment in the RAN and the core.

To initiate service outsourcing the custodian network would send an outsourcing request to the candidate network specifying parameters that will form a short-term SLA. If the SLA parameters are acceptable to the candidate network the user will be handed over to the candidate network. Session setup and eventual service delivery will proceed via the candidate network. If the outsourcing transaction is unsuccessful the call establishment attempt would be blocked by the custodian operator.

A service outsourcing agreement would be formulated using three main parameters:

- offer price (OP)
- outsource duration (T)
- required QoS resources (e.g., bandwidth)

The offer price represents the per unit monetary amount the custodian network would pay to the candidate network for the outsourced service. It would be a percentage of per unit payments the user would make to the custodian operator if service delivery occurred on the home network. The outsource duration represents the initial period of time an outsourced session would remain active on the candidate network before it could be re-called to the custodian network. The required QoS resources represent the QoS class and both uplink and downlink bandwidths required to transport user traffic during the duration of the outsourced session.

To maximise the benefits of service outsourcing the custodian operator would set lower offer price values to retain a bigger percentage of revenue accruing from user payments. On the other hand the candidate operator would be favoured by higher offer price values. The custodian and candidate operators may influence the proportion of shared revenue apportioned to them by manipulating the outsource parameters. It would be pointless for an operator to participate in service outsourcing if the cost of outsourcing was too high.

As a result of sharing revenue from user payments, custodian and candidate operators are faced with the possibility of lost potential income during the outsource period. The custodian operator would lose potential revenue if, during this period, network congestion subsided substantially to levels that avail capacity and resources large enough to accommodate the outsourced sessions. This would happen when some sessions were terminated by users or the network. On the other hand, the candidate operator would lose potential income if, after assigning resources to outsourced requests, service requests from their own home users could not be admitted to the network due to lack of sufficient resources.

It is essential for operators to evaluate the financial, short-term and long-term benefits of implementing service outsourcing. Various strategies may be adopted to influence the financial implications of outsourcing. Potential income losses may be minimised by negotiating favourable offer prices for a given outsource period. An additional measure for minimising potential income losses by the custodian operator may arise by prioritisation of user sessions for outsourcing. This would require service quality and cost differentiation on the custodian operator's network. This scheme may be complemented by scenarios where sessions may be handed over to cheaper access networks. Users of sessions that are charged at lower rates, and which require less stringent QoS guarantees may be marked as high priority for outsourcing to create capacity for higher profile sessions.

In addition to financial implications, service outsourcing would cause additional signalling in the network. This could further strain the scarce resources in the air interface¹. Thus, mechanisms are required to minimise signalling related to service outsourcing, while ensuring that effective outsourcing operations occur. Aggregation of outsourcing transactions may effectively contribute to minimisation of signalling. Outsource aggregation (or batched outsourcing) may be facilitated by prioritisation based on user service profiles. The use of profiles in 3G mobile communications was investigated in our previous work [3]. The use of profiles to support service outsourcing is explored in this work. We model a network where pricing and service cost, affect the acquisition of extra network capacity on short time frames in a multi-player environment.

¹It should be noted that congestion is the may trigger for outsourcing on the custodian operator's network

Multiple Candidate Networks

When multiple potential candidate networks exist the custodian network may send outsourcing requests to several networks, and select a target network after receiving and evaluating OP counter offers. By selecting a network with the lowest counter OP the custodian network would optimise on the monetary benefits of outsourcing.

In this thesis we investigate the required conditions for service outsourcing to be feasible and beneficial for involved operators. We explore the financial implications arising from involvement in service outsourcing for custodian and candidate operators. We propose and explore strategies that may be employed by operators to maximise the benefits of service outsourcing. Through analytical studies, mechanisms are developed for deriving optimal values of critical service outsourcing parameters. These mechanisms support the formulation of outsourcing strategies by operators. We investigate the practical use of service outsourcing in commercial network deployments. The IP Multimedia Subsystem (IMS) is used as a platform for service outsourcing. The IMS is a sub-system of the 3GPP packet switched (PS) core network that has been standardised for the provisioning of multimedia services using IP transport [4].

1.4 IP Multimedia Subsystem (IMS)

IMS is a standards-based IP-connectivity service control architecture for the delivery of various multimedia services to end users using common Internet protocols [5]. IMS is access-independent, hence different IP Connectivity Access Networks (IP-CANs) may be used for the interface between the User Equipment (UE) and the core network. The IMS architecture is divided into three planes: the media plane, the control plane and the application plane. In 3GPP networks IMS was standardised as a subsystem of the Packet Switched (PS) core network for the delivery of multimedia services like voice and video using IP. Examples of services and applications targeted for IMS are VoIP, IPTV using Video on Demand or broadcast methods and multimedia messaging. IMS call control is performed at the signalling and control plane, where the Call Session Control Functions (CSCFs) are located. Subscriber management is also handled at the control plane. Media traffic is confined to the media plane. Both media and control traffic

traverse respective IP-CANs. When an IP-CAN is congested, intended QoS levels for media traffic may not be achieved. A handover to an IP-CAN with adequate capacity to accommodate the media traffic would achieve the desired quality of communication.

IMS communications rely on standard protocols for information exchange between various entities. Session Initiation Protocol (SIP) [6] and Diameter [7] are the core protocols used in IMS transactions. SIP is a text-based protocol used for establishment and control of multimedia sessions. SIP is used between IMS clients and the IMS core, and between IMS core entities. Diameter is a secure Authentication Authorisation and Accounting (AAA) protocol used for authentication, policy and charging control.

Media traffic may traverse an IP core network on the end-to-end path between the communicating nodes. The providers of IP-CANs and IP core networks would charge for transport resources used to convey the traffic. Various business models that define accompanying charging mechanisms may be adopted for IMS communications. IMS charging may be separated and confined to individual planes. However, converged charging and billing may be adopted in scenarios where the IMS network and the IP-CANs are controlled by one operator. The Proxy CSCF (P-CSCF) is the entry point into the IMS core. The P-CSCF may be located in the IMS operator's home network or on a visited network in roaming scenarios. IMS call control is always performed by a registrar entity located in the home network. Thus IMS control traffic would traverse the home network. On the other hand, media traffic does not have to traverse the home network, e.g., in roaming scenarios with local breakout.

1.4.1 Service Outsourcing in IMS

According to standard IMS procedures an IMS client would register with the IMS domain after attaching to an IP-CAN. Signalling traffic is generated by registration events; however, the contribution to the aggregate network load by IMS registration events is not significant, especially when signalling compression is used. The IMS client would establish a media session by sending SIP INVITE requests to the IMS core. INVITE requests contain session description information of varying size; thus, the volume of consumed network resources may not be trivial. The actual media traffic, e.g., VoIP or video stream would consume more network resources, hence contributing significantly to network congestion. In a network that is operating at close to full capacity on a given IP-CAN IMS

clients may be able to register and send INVITE requests to the IMS core. However, lack of sufficient resources to transport media traffic may prevent establishment of media sessions. Service outsourcing may be used to facilitate service delivery via an IP-CAN belonging to a visited operator.

Signalling optimisation is required during service outsourcing to minimise the adverse effects of network congestion on the achievement of QoS guarantees for active flows. Transporting SIP messages over the air interface poses a challenge to the efficient utilisation of scarce resources. Service outsourcing in IMS frameworks will generate SIP signalling over the air interface. Excess SIP signalling during congestion periods will exert extra strain on meeting latency commitments leading to larger session setup delay.

Service outsourcing in IMS would use the 3GPP standardised S9 reference point [8] for inter-operator communication to exchange session description information and service outsourcing parameters. Diameter has been standardised for network AAA functions. In IMS service outsourcing, Diameter would be used to convey outsource parameters between custodian and candidate networks during the negotiation stage. Session description information contained in INVITE requests would contain requested bandwidth, which is required for service outsource negotiations.

The main idea behind service outsourcing is enabling real-time per session negotiation of shared resources. The cost of shared resources would be agreed before usage; subsequent charging for resource usage would occur during the lifetime of the session. In order to investigate service outsourcing in the IMS a charging framework is required. An open source IMS implementation is available for research; however, a charging framework is lacking.

In this thesis we develop and implement a standards compliant charging framework for the IMS. Open source IMS (OSIMS) was developed and released by the Fraunhofer Fokus institute in Germany [9]. OSIMS provides the standard IMS call session control functions, i.e., P-CSCF, I-CSCF, S-CSCF that are developed using C. They inter-work with an open source implementation of the home subscriber server (HSS).

1.4.2 Charging in IMS

Charging, billing and accounting are the most critical operational support activities for any telecommunications operator [10]. Charging is required for the network transport

service as well as for services offered on top of the operator's network. Transport services affect the IMS media plane, and require flow level charging. Flow level charging may be conducted on the home network or on the visited network in roaming scenarios. Service level charging involves IMS CSCFs and application servers. IMS supports offline and online charging at all planes. In offline charging, charging activities do not affect resource consumption in real-time. On the other hand, online charging affects resource consumption in real-time [11].

The charging framework developed in this project attempts to comply with standards. The main aim is to develop a functional prototype for use in research on aspects requiring charging, like service outsourcing. We develop a charging framework for flow level and service level charging of IMS communications. The testbed products include a charging application server and billing domain entities. Standard charging interfaces, i.e., Rf and Ro for offline and online charging are developed.

Service outsourcing requires inter-domain interactions between custodian and candidate networks. In roaming scenarios inter-domain charging transactions require an interface for the exchange of charging events and information. Interaction between IMS domains for the purpose of Policy and Charging Control (PCC) utilising standard reference points, e.g., S9 [8] have similar requirements with inter-domain service outsourcing interactions. Thus, we develop an interface to facilitate inter-domain interactions for service outsourcing. Service outsourcing parameters used in negotiation and maintenance of the session would be exchanged across the interface.

Policy Control in IMS

QoS policy control in IMS is achieved according to policy control standards from 3GPP, ITU-T etc. Initially policy control in the IMS was standardised for the UMTS access network, but later other access technologies were incorporated. The 3GPP QoS management architecture defines an Application Function (AF), a Policy and Charging Rules Function (PCRF), and a Policy and Charging Enforcement Function (PCEF). The AF would request resources from the transport plane; the PCRF performs resource management, while the PCEF resides on the transport (media) plane where network policies are enforced [12]. 3GPP policy and charging control are closely related; they use similar protocols and interfaces. For flow level charging control, Diameter based Rx and Gx

reference points are developed as part of the evaluation framework.

Signalling Protocols

IMS charging depends on the Diameter [7] protocol to convey charging events between Charging Trigger Functions (CTF) and either the Online Charging System (OCS) or the Offline Charging System (OFCS). Diameter transactions are mainly confined to the core network; hence these transactions do not contribute to the signalling overhead in the access network air interface. However, SIP signalling is used between the IMS user agents and the core. For service outsourcing transactions SIP messages will contribute to the signalling overhead. IMS Signalling optimisation is required to counter likely session setup delay in service outsourcing.

1.4.3 Service Outsourcing Signalling and Scalability

Signalling requirements for service outsourcing are affected by the number of messages exchanged between user equipments and custodian and candidate networks during the negotiation stage. SIP messages will be exchanged between base stations and user equipments on each network; thus consuming scarce air interface resources. Diameter messages would be confined to the core IP connection between the networks. To minimise the effects of excess signalling when processing many user sessions in peak periods, mechanisms aimed at signalling aggregation may be used to facilitate scalability of outsourcing transactions.

Aggregation of outsourcing transactions involves identifying sessions that may be outsourced, negotiating of outsource requirements for multiple sessions with candidate networks, and signalling target UEs using a broadcast message to handover to assigned candidate networks. The batch processing of outsourcing transactions may be combined with pre-emptive outsourcing measures. Pre-emptive outsourcing involves prediction of imminent outsourcing based on congestion thresholds. Pre-emptive outsourcing may be complemented by the use of service profiles to prioritise sessions to be outsourced. The QoS requirements of active sessions may also be considered in prioritisation of sessions for pre-emptive outsourcing. Lower profile sessions may be outsourced to free resources for expected higher profile sessions that would generate more income for the custodian network.

Quality of Service and Pricing Profiles

Service profiles may be used to classify services as premium, or basic. Network operators may create service profiles that map QoS guarantee levels to the cost of service. Higher QoS profiles would attract higher usage fees. In heterogeneous environments these service profiles may be mapped to IP-CANs that meet the QoS characteristics defined by the profile. Custodian operators may place a higher outsource priority to users of lower QoS profiles as network congestion levels rise. Sessions belonging to these users may be pre-emptively outsourced when service requests from higher profile users are received.

1.5 Service Outsourcing Benefits and Challenges

The benefits of service outsourcing revolve around sharing network capacities to meet peak resource demands. The custodian network will virtually expand its capacity to meet peak period service requests; this is achieved without expensive infrastructural deployments. On the other hand, the candidate network is able to utilise extra capacity, thus improving efficiency and sharing income with the custodian operator.

Service outsourcing requires operators to share income accruing from user payments for service usage. This is due to the need to maintain the same service fees for users regardless of the network to which they are attached. Networks participating in service outsourcing will experience reduced potential net income. Moreover, service outsourcing income will vary with the number of sessions, offer price values and outsource duration. The custodian network would have earned full income if there had been sufficient resources to accommodate all user service requests. On the other hand, the candidate network would earn full income from its resource capacity if all actively connected users were its subscribers.

The short-term service outsourcing SLA defines an outsource duration over which the session remains active on the candidate operator network. During this period the custodian operator may lose potential income while having unused resources on the network. The candidate operator earns less income when home sessions get blocked due to assigning resources to outsourced sessions. This may be compounded by dissatisfaction of blocked candidate network users. Thus it is essential to analyse the financial implications of service outsourcing for custodian and candidate networks, and formulate algorithms

aimed at maximisation of gains from outsourcing.

1.6 Thesis Objectives

This thesis will address the concept of service outsourcing as a mechanism for facilitating real-time resource sharing amongst competing mobile networks. The research will be illustrated using IMS communications.

The thesis will explore the following items:

- Conduct a review of literature on resource sharing solutions for mobile and wireless networks.
- Highlight and compare the benefits of different resource sharing solutions to clearly identify models that would be beneficial to participating operators.
- Review literature on 3GPP standardised IP Multimedia Subsystem communications. Emphasis will be placed on IMS charging procedures and session control with QoS.
- Discuss the concept of service outsourcing and formulate generic service outsourcing procedures and contracts.
- Explore the technical conditions and requirements of service outsourcing involving only two operators, and eventually consider multi-operator scenarios. The technical conditions would revolve around resource requirements for services requested by users, and resource capacities of custodian and candidate networks to meet the requirements.
- Explore the benefits of service outsourcing for custodian and candidate operators. Financial and other outsourcing benefits will be discussed.
- Analyse the financial implications of service outsourcing. Analytical models will be developed based on the terms of outsourcing contracts. Algorithms encompassing strategies to be applied by operators to maximise on benefits will be formulated.
- Identify service outsourcing challenges with regard to signalling overhead and formulate mechanisms for signalling optimisation.

- Discuss service outsourcing with respect to IMS communications; the effects of service outsourcing on IMS session control will be explored.
- Design and develop a charging prototype for IMS communications research. IMS service outsourcing tests will be performed on an IMS framework with charging enabled.
- Perform proof of concept tests involving IMS service outsourcing between two networks, and collect performance results for analysis. The results will validate the feasibility of implementing service outsourcing for multimedia communications.

1.7 Thesis Scope and Limitations

Service outsourcing is a form of resource sharing amongst mobile networks. Resource sharing involves many aspects of call session control, QoS negotiation, SLA establishment between operators, SLA conformance monitoring etc. Call control for active IP sessions requires mobility management schemes like mobile IP.

Dimensioning of network resources to meet user service demands will be abstracted at the IP level. Service outsourcing would involve handovers and mobility between networks. Various mobility management schemes, e.g., mobile IP have been developed for IP communications. Discussions of such schemes will be confined to the literature review, since the service outsourcing scenarios explored in this thesis will adopt a higher layer session mobility scheme, e.g., SIP based schemes.

Service outsourcing may be applied to different network architectures and access networks using different technologies. This thesis will focus on 3GPP standardised networks, i.e., 3G and the Long Term Evolution (LTE). The evolution towards LTE creates networks with sufficient capacities to support multiple multimedia sessions with performance guarantees. We explore service outsourcing in the framework of IMS communications. A framework independent discussion of service outsourcing will be presented. However, discussions involving signalling protocols will be based on IMS communications. Moreover, the charging prototype will be IMS specific.

Analysis of the numerical results will be conducted for a generic service outsourcing scenario to support conceptualisation of service outsourcing and its benefits. Testbed results will be confined to proof of concept and performance evaluation tests.

1.8 Thesis Contributions

The main contributions of this thesis are:

- An extensive review of network sharing solutions used by mobile operators to lower network deployment and operation costs. These sharing solutions form the greater sphere of network resource sharing to which service outsourcing belongs. We introduce, conceptualise and formulate mechanisms for service outsourcing. We create definitions for this resource sharing scheme to facilitate further research in this area.
- Analysis and design of a service outsourcing framework for wireless communications. We analyse options available to operators and formulate procedures for sharing resources to meet peak service demands, sustain subscriber satisfaction and avoid expensive network upgrades.
- Expand the functionality of the UCT IMS client [13], an Open Source IMS Client implemented in C for the Linux platform. The functionality added supports user interaction with the charging and billing system to select billing account options and receive advice of charge.
- Author of the UCT Charging Control Framework, a 3GPP standards compliant free Open source IMS charging system. The charging system supports online and offline reference points based on the Diameter protocol. It utilises the Open Source IMS core system released by the Fraunhofer Fokus institute [9].
- Design and implementation of a inter-domain resource negotiation and service bidding framework for IMS networks. This system enables different IMS operators to dynamically adjust prices of their network resources based on demand for resources, and choose to participate in service outsourcing.

The following publications document the contributions of this thesis:

Journal:

1. Vitalis G. Ozianyi, Neco Ventura, Eugene Golovins, "A Novel pricing approach to support QoS in 3G networks," Elsevier computer networks, vol. 52/7 pp 1433-1450, May 2008 [14].

Conferences:

1. Vitalis Ozianyi, Neco Ventura, "Real-time Resource Sharing Through Service Outsourcing", Proceedings of the Local Computer Networks International workshop on the Performance and Management of Wireless and Mobile Networks (P2MNET), Denver, USA, October 2010 [15].
2. Joyce Mwangama, Vitalis Ozianyi, Neco Ventura, "Charging and billing for composite services in a multi-service provider environment: the IMS case," Proceedings of the IEEE Wireless Communication Networks Conference (WCNC), Sydney, Australia, April 2010 [16].
3. Vitalis G. Ozianyi, Neco Ventura, "Design and Implementation of Scalable IMS Charging Systems," Proceedings of the Local Computer Networks International workshop on the Performance and Management of Wireless and Mobile Networks (P2MNET), Zurich, Switzerland, October 2009 [17].
4. Vitalis G. Ozianyi, Neco Ventura, "Efficient and scalable charging systems for IP multimedia networks," IEEE Africon conference, Nairobi, Kenya, September 2009 [18].
5. Vitalis Ozianyi, Richard Good, Phillippa Wilson, Neco Ventura, "Optimized IP-CANs to support best charged IMS scenarios," IEEE International symposium on personal indoor and mobile radio communications (PIMRC 2009), Tokyo, Japan, September 2009 [19].
6. Vitalis Ozianyi, Neco Ventura, "Service outsourcing and billing in inter-domain IMS scenarios," South African Telecommunications Networks and Applications Conference (SATNAC 2009), Mbabane, Swaziland, September 2009 [20].

7. V. Ozianyi, R. Good, N. Carrilho, N. Ventura. "An XML driven framework for policy based QoS management in IMS networks." Proceedings of IEEE Global communications conference (Globecom), New Orleans, LA, USA, Nov-Dec 2008 [21].
8. Vitalis G. Ozianyi, Neco Ventura, "Service outsourcing to support virtual home environments (VHE)," Proceedings of the Local Computer Networks International workshop on the Performance and Management of Wireless and Mobile Networks (P2MNET), Montreal, QC, Canada, October 2008 [22].
9. Vitalis G. Ozianyi, Vitor Jesus, Susana Sargento, Rui Aguiar, Neco Ventura, "Virtual network capacity expansion through service outsourcing." Proceedings of the IEEE Wireless Communication Networks Conference (WCNC), Las Vegas, NV, USA, Mar-Apr, 2008 [1].

1.9 Thesis Outline

The rest of this thesis is structured as follows:

Chapter 2 presents an extensive literature review of resource sharing in wireless networks. Issues regarding spectrum sharing, RAN sharing and CN sharing are covered here. This chapter also reviews literature on mobility management including roaming in wireless networks. An extensive review of literature on the use of SLAs and different business models involving communication networks is presented, as well as discussions on Internet economics with regard to the supply and demand of resources.

Chapter 3 presents a comprehensive review of standardisation work and literature on IMS communications. Primary focus is given to charging control standardisation by 3GPP. This chapter contains foundation information supporting discussions in subsequent chapters.

Chapter 4 models and presents the service outsourcing concept. Service outsourcing is approached as a real-time network resource brokering mechanism. We analyse the utilisation of network resource capacities and create algorithms for resource sharing. We introduce critical parameters required in formulating SLAs for short-term resource sharing, and highlight strategies that may be employed by networks to maximise the

benefits of service outsourcing. The design of an inter-domain resource broker is presented to support service outsourcing. Using a generic scenario, we demonstrate service outsourcing transactions to give a clear view of operator benefits.

We present service outsourcing in IMS in chapter 5. A charging framework is required to completely illustrate service outsourcing in IMS. The complete design of a standards compliant charging framework for IMS is presented. The design supports charging at the media, control and service planes. We present additional mechanisms required to support real-time resource brokering and negotiations between networks.

Chapter 6 presents a testbed implementation of service outsourcing in IMS frameworks. The implementation includes an IMS standards compliant charging framework with resource brokering capabilities. The testbed is clearly structured and documented to permit extensibility to support further research involving charging and resource sharing.

In chapter 7 we perform extensive tests on the testbed. Both proof of concept and performance evaluation tests are conducted and results are presented.

Chapter 8 presents conclusions and final discussions. A summary of the research is presented, as well as recommendations for further research in this area.

Chapter 2

Literature Review

2.1 Introduction

In this chapter we review trends adopted by mobile network operators in an attempt to meet the continuously increasing demand for network resources. This includes long term options like capacity planning and network upgrades, as well as short term options like network sharing. This chapter provides a motivation for the work done in this thesis. The information presented provides a base for the discussions and models presented in chapter 4.

Network resource sharing can enable operators to save on CAPEX and OPEX. They contribute to the final charge levied to subscribers for using telecommunication services. If these costs are lowered the ultimate charge to users for service usage will likely be lower. In providing an end-to-end solution, the price charged by different networks and domains along the end-to-end path between communicating nodes contributes to the total service cost towards end users. End users may connect through different access networks to core networks that are interconnected with other networks to form end-to-end paths. In a QoS-enabled DiffServ IP network, resource provisioning to service classes in core networks is guided by service level agreements. Resource provisioning in the IP core is governed by long time frame SLAs; resources are eventually sold or allocated to users in access networks on short time frames. This is the case where ISPs enter into long term SLAs with higher tier providers, but allocate resources on demand basis to their subscribers. Capacity provisioning and management through pricing may

be achieved by using wholesale prices for the aggregate resource blocks in the IP core and retail prices for resource allocations to end users on access networks.

DiffServ [23] is designed to allocate network bandwidth and queue priorities to network traffic that is grouped into classes. Admission control and resource allocation mechanisms are needed in DiffServ to prevent severe degradation of performance for flows in a given class under congestion [24]. The use of pricing and dynamic tariff mechanisms to control congestion has been thoroughly investigated [25]. Such pricing methodologies suggest raising prices with increasing congestion in order to deter price sensitive users from joining the network during peak periods. The effectiveness of such mechanisms is limited to how high prices may be increased without impacting greatly on user QoE, and remaining competitive in an environment with multiple network operators.

Communication networks have limited resource capacity, thus as users get admitted to the network the available capacity gets consumed. Generally networks are designed to meet average peak resource demands; thus in peak periods the capacity is insufficient to meet the actual demand. The behaviour of different applications varies under network congestion conditions. For instance the performance of real-time multimedia applications is adversely affected when congestion leads to intolerable packet delay, jitter and excessive packet drops. Generally congestion results in poor service delivery. Fresh capacity planning involving network upgrades is required for provisioning more resources to avert recurrent congestion as well as meet increased demand.

Upgrades to mobile networks require an end-to-end solution involving wireless access networks, backhaul and the core network. There has been steady evolution of networks aimed at increasing the amount of resources, e.g. bandwidth, available to user applications. The evolution of communication networks is a continuous process whose sustainability requires eventual recouping of capital investments within the lifetime of a given technology. Operators aim to provide services to their subscribers in a well defined and controlled manner, and at costs that are competitive and affordable to users. Profit maximisation is an important goal of each operator.

Network business models developed to facilitate QoS provisioning require schemes for dimensioning of networks to avail adequate resources to applications and prevent congestion. The DiffServ architecture is inherently a business model focusing on correctly meeting the supply of resources by sizing each logical network [26]. Demand for resources is more volatile, hence hard to predict. SLAs have been designed to assist

network providers in predicting the expected demand for resources over a given period. However, the use of SLAs to constrain demand for resources is not effective in avoiding occasional congestion when the unpredictable demands from many customers coincide. Moreover, SLAs are non-transitive; this means that performance guarantees are defined between peer domains and are confined to individual provider domains. It becomes challenging to extend SLAs defined for the core network to cover the access network. This is pronounced if the access link and core network used by media traffic belong to different providers. Wireless access links are inherently lower bandwidth connections that are prone to congestion.

Upgrading networks with new technologies is the common method of dealing with increasing demand for resources. Wireless networks have been experiencing steady evolution from low bit-rate connections to high bandwidth, low latency broadband connections. Multimedia service control frameworks are developed with the assumption of over-provisioned core networks; hence the QoS constraint is considered to exist in access networks.

2.2 Next Generation Wireless Access Networks

The evolution of wireless access networks is a continuous process aimed at meeting the needs of new applications that result in the constant growth of network traffic. Over time different mobile radio access technologies have been developed and deployed. The 3GPP has been responsible for standardising wireless networks in the family of Global Systems for Mobile communications (GSM). The GSM family of networks was initially standardised by the European Telecommunications Standards Institute (ETSI), and later by the 3rd Generation Partnership Project [27].

The first generation of GSM networks was designed for voice traffic only. Low bit-rate data services were introduced in 2nd generation networks. Circuit Switched Data (CSD) and High Speed CSD (HSCSD) supported data rates of 14.4kbps and 57.6kbps respectively. Packet data transport was introduced in 2G networks with the General Packet Radio Service (GPRS) providing about 144kbps. This was later enhanced to EDGE (Enhanced Data rates for GSM Evolution) providing data rates of about 384kbps over the GERAN [28]. Table. 2.1 illustrates the evolution of data services in the GSM family of networks. These early packet data services used the same frequency spectrum

(900 and 1800 MHz bands) as GSM voice services of the same generation. The data rates were too low to support multimedia services.

Table 2.1: The evolution of GSM wireless networks

Network Generation	Technology	Data rates	Multimedia support	NGN
2	GSM, HSCSD, GPRS	144kb/s	No	No
	EDGE	384kb/s	No	No
3	UMTS	2Mb/s	No	Yes
	HSDPA, HSUPA	14Mb/s	Yes	Yes
	HSPA	11.5 - 28.8 Mb/s	Yes	Yes
4	LTE	50-100 Mb/s	Yes	Yes

The evolution towards 3G networks or the Universal Mobile Telecommunication System (UMTS) was aimed at increasing the bandwidth available in the air interface, as well as achieving a separation between service and transport planes [27]. The achieved Next Generation Network (NGN) evolved to incorporate the IP Multimedia subsystem (IMS) in the packet switched core network in release 5 of 3GPP standards. Evolution to an all-IP UMTS network followed with the introduction of IP transport within the UTRAN. The High Speed Downlink Packet Access (HSDPA) was introduced in 3GPP release 5 as well. It provided improvements in downlink data rates and capacity. The High Speed Uplink Packet Access (HSUPA) was standardised in release 6 to improve capacity and data rates in the uplink direction. In release 7 the High Speed Packet Access (HSPA) was introduced. The data rates specified for HSPA are 28.8Mbps and 11.5Mbps in the downlink and uplink directions respectively [29]. These data rates are achieved using Multiple Input Multiple Output (MIMO) antenna solutions and higher order modulation. Moreover, advanced mobile receivers can sustain higher data rates in adverse channel conditions. HSDPA, HSUPA and HSPA data rates are sufficient to support different types of mobile multimedia applications, e.g., multiple VoIP sessions and IPTV.

HSPA adopts a flat network architecture that reduces the number of network nodes on the user plane to 2, i.e., an eNodeB and an Access gateway. The flat architecture complies with 3GPP release 8 standards for the Evolved Packet System (EPS), incorporating Long Term Evolution (LTE) and the Evolved Packet Core (EPC). A simplified illustration of the EPS is shown in Fig. 2.1. Appendix A.1.3 presents more information on LTE. Thus

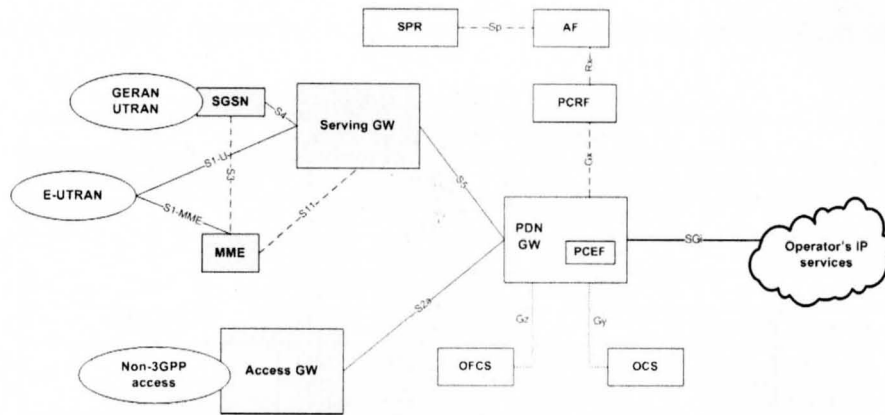


Figure 2.1: A simplified version of the evolved packet system

HSPA was viewed as a pathway to LTE. LTE data rates will exceed those of HSPA. The expected bandwidth for the uplink and downlink connections will be approximately 50Mbps and 100Mbps respectively [30]. Radio frequency spectrum utilisation influences the evolution of radio access networks. New generations of cellular networks have been designed to efficiently utilise scarce radio resources. The introduction of Femtocells is an attempt at utilising very small cell sizes for cellular access. Femtocells may be the only feasible cellular option for reaching areas where macro-cell signal quality is poor due to rapid attenuation.

3GPP has standardised the use of femtocell devices and systems for UMTS (UTRAN) family of networks. Femtocells are aimed at improving capacity and coverage in indoor areas, homes and public hotspots. The business case surrounding the deployment of femtocells involves the use of existing fixed broadband technology infrastructure like xDSL lines for backhaul [31]. Operators can save on OPEX and CAPEX while effectively utilising purchased licenced spectrum.

In addition to wide-area mobile cellular technologies, Wireless LANs (WLAN) have been deployed in hotspot areas to provide data transport. Popular WLAN technologies have been standardised by the Institute of Electrical and Electronic Engineers (IEEE). These include IEEE 802.11a that operates in the 5GHz band with an effective bandwidth of 30Mbps, IEEE 802.11b operating in the 2.4GHz band with an actual throughput of about 5.5.Mbps and IEEE 802.11g [27]. Table 2.3 illustrates different types of wireless local area networks. To support QoS in WLANs the IEEE 802.11e standard was introduced. Interworking between these WLANs and 3GPP systems was standardised in release 6 of

3GPP standards [32]. Appendix A.1.1 presents supplementary information on WLAN.

Table 2.3: Wireless local area networks

IEEE Wireless Networks			
Network Name	Technology	Data rates	QoS Support
Wi-Fi	IEEE 802.11a	54Mb/s	No
	IEEE 802.11b	11Mb/s	No
	IEEE 802.11g	54Mb/s	No
	IEEE 802.11e	-	Yes - 4 classes
Wi-Max	IEEE 802.16d	70Mb/s	Yes - 4 classes
	IEEE 802.16e	70Mb/s	Yes - 4 classes

In addition to WLANs IEEE created a standard for wireless metropolitan area networks, the IEEE 802.16 standard for WiMax [33]. Wi-Max is standardised to provide broadband wireless access with QoS to users. The standard specifies physical transmission using OFDM and a connection oriented MAC layer. The initial frequency spectrum range for WiMax was 10-66 GHz for the IEEE 802.16 Line-of-Sight (LoS) dependent transmission. The 2-11 GHz band was allocated for IEEE 802.11a-2003 standard, which supported non-LoS transmission as well. IEEE 802.16d was released in 2004 with several enhancements to previous releases to address first or last mile connections for WMANs. Mobile WiMax was introduced in the IEEE 802.16e standard. The standard uses antenna diversity and other schemes e.g., MIMO to improve non-LOS coverage and indoor penetration. The theoretical maximum bandwidth of WiMax is about 120Mbps with an expected maximum transmission range of 50 to 70 km. Appendix A.1.2 addresses this technology by providing additional information.

Network operators may deploy various access networks to support different business models surrounding Internet access and telecommunication services. Providing effective data transport and media rich services and applications in addition to traditional telecommunication services has been a driving factor for the deployment of access networks based on different technologies.

Heterogeneous access networks are an important feature of NGN. Service outsourcing will be integrated in the heterogeneous environment to exploit the properties of different access technologies. From the operator's perspective the cost of providing services depends on the access technology. This will influence network decisions on service outsourcing transactions, especially when maximisation of benefits could arise from using a specific access network.

2.3 Network Sharing Solutions

Upgrading of networks to meet increasing demand for resources is only feasible as a long term solution to the imbalance between supply and demand of network resources. Network sharing can be used as a short term as well as a long term solution to improve network coverage and capacity. Network sharing solutions structured on long term agreements to allow sharing of operator infrastructure include: site sharing, radio access network sharing and core network sharing. Network roaming is a network sharing solution based on long term business agreements, but resource sharing occurs over short time frames.

2.3.1 Site Sharing

Site sharing allows two or more operators to share network facilities like cellular masts, power sources etc [2]. Site sharing is feasible in areas with high business potential and characterised by heavy competition between operators [34]. Each operator has full control of their network assets; this includes dedicated frequencies and core networks. Fig. 2.2 illustrates site sharing in a typical 3G network [34]. The shared site is either owned by operator 1 or operator 2, or ownership can be a joint venture of the operators or a 3rd party provider. CAPEX and OPEX savings are directly proportional to the number of shared facilities, i.e., site space, power sources, antenna radomes etc.

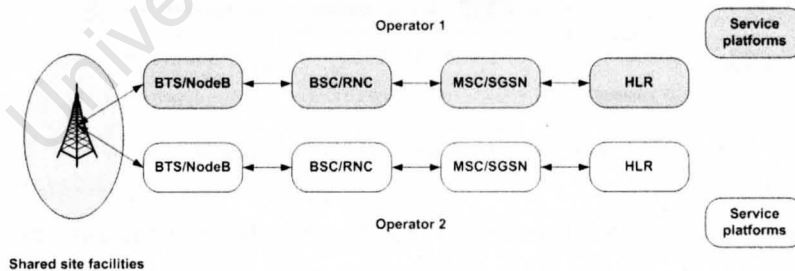


Figure 2.2: Site based sharing solution

2.3.2 Radio Access Network (RAN) Sharing

A RAN comprises the set of network components that enable a user equipment to interact with the network using radio frequency waves. In a Universal Mobile Telecom-

munications System (UMTS) network the RAN includes the NodeBs, the Radio Network Controllers (RNC) and links between them. RAN sharing is feasible in areas with moderate business potential and normal competition between operators. Each operator retains partial control of network assets. In a typical 3G deployment RAN sharing has the following characteristics: dedicated or shared radio frequencies, shared site facilities, shared BTS/NodeB and shared BSC/RNC [2]. Fig. 2.3 illustrates RAN sharing for a 3G network [34].

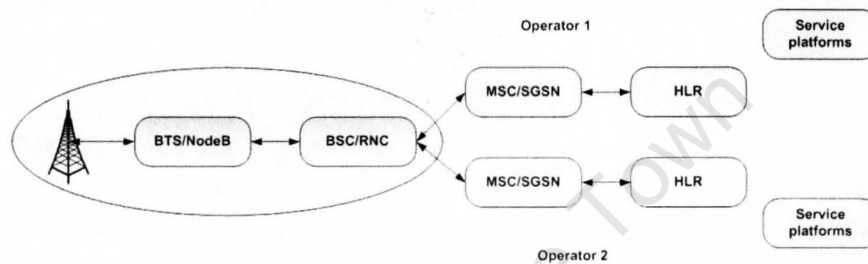


Figure 2.3: Radio Access Network Sharing

2.3.3 Full Network Sharing

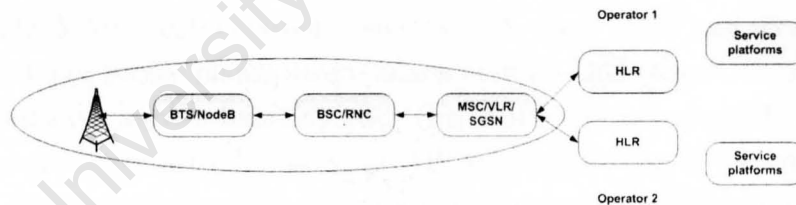


Figure 2.4: Full network sharing

Full network sharing involves sharing of RAN and core network equipment [2]. It is feasible in areas with low business potential, and requiring coverage possibly due to regulatory requirements. In full network sharing operators share all network entities except the subscriber databases, i.e. HLR or HSS, and service platforms. Fig. 2.4 illustrates full network sharing [34]. The shared network can be owned by either or both operators, or a third party.

Network costs account for 33% of the total operator expenditure [34]. Network sharing

solutions enable operators to save on network costs. The network sharing solutions discussed above are formulated to minimise costs related to improving network coverage. Moreover, they are based on long term agreements that are not flexible in terms of service charges and adapting to short-term changes in the demand for network resources. Mobile Virtual Network Operators (MVNO) exploit full network sharing by making service agreements with numerous network operators [35]. To optimise network costs and meet resource demand from users the MVNO can select amongst available networks. Users can be served on an alternative network if they cannot be admitted to a given network due to congestion.

Mobile Virtual Network Operators

Operators are faced with multiple challenges related to investing in technologies to provide services to users. The main challenges include: raising the finance to pay for frequency spectrum, technology and development; choosing and implementing the right technology; commercialising the technologies and business development [35]. With respect to new technologies, operators are faced with the dilemma of whether to buy or build new networks in order to benefit from arising income opportunities. MVNO provide services on top of various mobile infrastructure operators (MNO) by buying resource capacities. The MVNO enters into agreements with many MNO and service providers to provide end-user profile management, access network QoS control, administration of residential gateways, service bundling etc. This avails a broad range of enhanced customisable services to users. To users the MVNO appears like a conventional MNO. MVNO can offer expanded service choices and cost reduction whose extent depends on the terms of payment to the MNO. MVNOs play a vital role in the provision of Virtual Home Environments to mobile services subscribers.

Virtual Home Environment

In a virtual home environment (VHE) end users can access the services available on their home networks when roaming on other networks. VHE allows a user to personalise the set of services for which he has a subscription with the home network, and provides these services with the user's personalised "look and feel" across different types of networks [36, 37]. Heterogeneous NGNs enable users to access services using different Radio Access

Technologies (RAT). Moreover, users should be able to choose a network or an access technology that gives them the best connection and service experience [38]. In order to support inter-operation amongst different networks, access network coupling and service coupling is required to facilitate service provision [35]. Network coupling is needed to support mobility across networks using the same RAT (horizontal handoff) and also those using different RAT (vertical handoff) [35]. SLAs are needed to support roaming between and amongst different operator networks. Bilateral agreements can be used, or a third party, like a Network Inter-operation Agent [39], can be used to facilitate AAA interactions amongst operators.

With regard to service provision, different service coupling strategies can be used depending on service requirements to facilitate a VHE. However, the generic service provisioning strategy is to have an interface linking the network technology specific service control layer with the network independent application layer, thus allowing applications to discover the capabilities of the underlying network. An example framework is based on the Open Services Architecture (OSA), and can be mapped onto telecom specific protocols like the Customised Applications for Mobile Enhanced Logic (CAMEL) Application Part (CAP) and the Mobile Application Part (MAP), to provide listener-controller capabilities for different services. Extensions to the OSA framework were proposed in [35] to add network re-configurability for QoS, user customisation, and support of adaptable service provision and flexible charging schemes.

Network Roaming

Roaming is the ability of a user to access wireless telecommunication services in areas other than where their subscription resides. Roaming can be implemented by means of a bilateral agreement between operators, or a roaming broker to allow users to access widespread mobility through providers with roaming contracts with the broker. The broker provides signalling support and charging information [35]. Roaming outside the coverage of a home operator's network involves attaching to visited operator networks. The UE discovers available networks based on information broadcast by local networks. Using the discovered networks' IDs, the user can manually select a network to attach to, or automatic network selection can be achieved using a preference network list stored on the USIM [32].

A roaming user session can be subjected to home domain control, visited domain control

or control procedures can be distributed among both domains. The devices interact directly with home servers in the home domain control case. The home domain retains the service profile and service related data, and core logic. On the other hand, in the visited domain control scenario the visited network downloads and temporarily stores this information. Resource management, accounting and charging are primarily performed by the visited network. The home network also performs accounting and charging for all user events and sessions. Generally the roaming user is charged higher service usage fees as compared to services offered over the home network.

Roaming facilitates network sharing primarily by providing access coverage to roaming users. User traffic and service control transactions traverse the visited domains's RAN. Depending on the service control scenario, user and control traffic can be routed to the home domain; alternatively local breakout for user traffic can occur in the visited domain. In the later case only control traffic is routed to the home domain. Mobile IP has been developed as the protocol to achieve host mobility in IP networks. Additional protocols have been standardised to support AAA functions.

Mobility among different networks involves performing handoffs using schemes designed for each access technology. Horizontal handoff for example intra-subnet mobility in IEEE 802.11WLANs can be facilitated by a layer 2 distribution network. It performs an update of the association between a WLAN station's MAC address and the serving AP. On the other hand vertical hand require co-ordination between the technologies used by each network. This can be achieved using a loosely or tightly couple approaches. Mobility in WiMax has been extensively developed for IEEE 802.16e networks. Mobile WiMax supports various forms of handoff i.e. hard handoff, Fast Base Station Switching and Macrodiversity Handover [33].

Inter 3G-WLAN roaming can be achieved with inter-connected AAA systems. In this case the roaming WLAN user's subscription information is stored in the Home Subscriber Servers (HSS) of home operator. 3GPP standardised mechanisms for visited network selection may be used in WLAN roaming. In practise all WLAN UE traffic can be tunnelled to the 3GPP core network Packet Data Gateway (PDG).

The network sharing solutions discussed above, including the existence of MVNO and roaming use long terms agreements covering the services users may access on visited networks. The agreements also cover the cost of service and settlements between home

and visited operators for aggregated user activities. The main motivation behind the network sharing solutions is in providing extended network coverage in unreached zones.

Network operators, both MVNOs and MNOs need mechanisms for rapid establishment of short term SLAs for network sharing. The SLAs should cater for determination of the cost of resources (service) based on current network utilisation levels, and allow for renegotiation of SLA terms as network conditions change. Moreover, performing SLA negotiation on a per session basis creates sufficient motivations for operators to engage in network sharing to virtually increase the capacity of their networks during congestion periods. This implies that operators can still engage in network sharing in zones where they have sufficient coverage, e.g., areas with high business potential and heavy competition among operators.

In this thesis a network sharing solution is formulated based on short term SLAs negotiated for each session to achieve virtual network capacity expansion in zones with wireless overlaps between operators. The solution, moreover, shields users from price variations like those associated with roaming.

2.4 Service Level Agreements (SLA)

A service level agreement is a contract between a network provider and a customer defining all aspects of the service to be provided [40, 41]. SLAs regulate the relationship between operators, ISPs and users. They are generally static contracts drawn periodically, e.g., monthly, and cover availability, performance and customer service. Consumers require performance guarantees on the end-to-end transport path between communicating network nodes. This includes performance assurances across the access link. SLAs provide service differentiation, and gain from economic benefits of service differentiation. Moreover, SLAs enable users to choose services that are bundled with the features they require, and priced accordingly.

An SLA must specify the general contractual conditions, i.e., duration, time, payment; the QoS performance in terms of service level specifications, i.e. delay, jitter, packet loss, the traffic profile and traffic conditioning functions [24]. SLA rules involve traffic description, geographical scope, duration and cost. DiffServ supports SLA negotiation between users and the network. More information on DiffServ is given in appendix A.2.

It is very challenging to define and meet SLAs for Internet communications due to the inherent best effort style of operation. Performance guarantees defined in an SLA may not be adequate in meeting the requirements of an application, e.g., VoIP. This is particularly the case when cross-domain communication is involved. SLAs are non-transitive across domains; moreover service-level parameters for an SLA class in a given domain may not map to a similar class in an adjacent domain [41]. Various solutions have been proposed to extend SLAs beyond peer domains: the solutions include: least-effort policy, most-effort policy and equal distribution policy [41]. The solutions define the extent to which each transit domain goes to meet the service-level requirements of an end-to-end request. The main metrics used in formulating SLAs for IP networks are bandwidth, delay, jitter and packet loss.

An additional metric that is critical to sustaining good QoE for users is network availability. This relates the network up time over a given period. It should be noted that a network may be available but not usable, e.g., as a consequence of congestion. The use of long time frame SLAs are aimed at protecting the service provider; thus the user is not offered significant assurance levels [40]. Monitoring of adherence to the terms of an SLA is required to facilitate resolution of customer complaints when a violation occurs; settlements of SLA violations are required.

The performance of an IP network may be determined in several ways. The use of probe packets, e.g., periodic ping requests is very common. Alternatively, measurement based schemes may be used, but these are resource intensive, thus not popular. Application level performance metrics are required to get a better assessment of the performance of a network. For web traffic, web response time is an application level metric that accounts for the user's perception of network performance. It combines network delay and packet losses in the performance evaluation; users will get frustrated when pages take long to load. A response time of more than 10 seconds is considered to be too long [40].

Service outsourcing depends on SLAs to govern the delivery of services to users of outsourced sessions. The decision of values to be used for outsourcing parameters will go into the formulation of an SLA between the operators. The SLA is important for the success of service outsourcing, and settlement of bills between the operators. A home operator has an implicit SLA with subscribers to provide services around the clock. It is for this reason that service outsourcing enables operators to overcome inadequate network capacity that would hinder meeting SLA conditions.

2.5 Internet Economics

The relationship between the supply and demand of network resources, and its influence on pricing of network services has been investigated extensively. Operators study resource demand patterns over daytime and nighttime. Moreover, operators try to influence demand patterns using pricing as an incentive. Pricing of network services has been associated with different access networks in developing business models for services customised for the access networks. The popularity of VoIP and other multimedia services has been boosted by pricing schemes that appeal to users and suit operator interests in terms of network operation and management. Generally pricing schemes are chosen to influence resource usage and maximise operator profits.

2.5.1 Pricing and Charging Schemes

Various pricing schemes have been proposed, some have become popular with Internet communications, whereas some remain popular with traditional telephony.

Flat-rate Pricing

Flat-rate pricing is the most popular charging scheme in the Internet. Users are charged a fixed subscription fee for a specified period, e.g., one month. Resource usage accounting information is not required to facilitate this scheme; however, factors like bandwidth of the access connection are taken into account to formulate the subscription tariff. Moreover, a data cap may be used to limit usage level for high speed connections.

Volume Based Pricing

Volume dependent charging accounts for the transmitted number of bytes or packets of information. Volume based charging is required to constrain resource usage, for example in wireless access networks, where the cost of provisioning is high. It is a key component of charging schemes aimed at controlling or curbing network congestion. In general terms the volume based tariff may be expressed as: $C_V = c_v V$, where c_v represents the charge per unit (e.g. charge per-byte), C_v is the computed cost and V is the volume of traffic. This usage charge is normally set to allow a retail network to recover the cost of purchasing resources from a wholesale network, plus other fixed costs.

Duration Based Charging

This charging scheme considers the connection time in determining network usage levels. Duration charging only works when resource demand is near constant across the period of interest. This charging scheme is popular with circuit switched services. It has been adopted for many session based packet services that traditionally were provisioned using circuit switching. VoIP is an example of such services.

QoS Dependent Pricing

This pricing scheme introduces a congestion-dependent component to provide an incentive for adaptive applications or users to change their service class or sending rates according to network conditions [25]. QoS pricing accounts for network performance by applying higher charges for improved performance. User willingness to pay extra for improved performance during congestion periods will depend on their utility function, and demand curve [24]. Network congestion can depend on the time-of-day, hence ToD pricing is common in telecommunication networks. In this case pricing is used to control congestion by exploiting the sensitivity of some users to price variations. User attitude to network prices is used to facilitate admission control and avert congestion. Congestion may be avoided by exploiting the potential for users to adapt their demand according to changing prices.

When using pricing to control congestion it is important to use prices that maximise profits and minimise call blocking or packet dropping at the same time. Call blocking at the admission control level will frustrate potential users. Various proposals for QoS pricing: responsive and priority pricing. In the latter, the price dynamically changes with network congestion as notified by feedback mechanisms, whereas in the former packets are charged according to the class they belong. Dynamic pricing requires fast reaction to changing network conditions. Human intervention is not fast enough to respond to network triggers like Explicit Congestion Notification (ECN). Middleware can be used to solve the fast control problem and overcome short-term network congestion [26]. This approach, however, results in increasing service costs for users who need high QoS during the congestion periods.

QoS dependent prices can be formulated by considering the supply and demand of network resources over a given duration. Bandwidth and buffer space can be allocated to

meet demand from traffic flows. Congestion prices can be re-adjusted if demand exceeds the allocated resources in a given time frame.

Inter-domain Pricing

In Internet communications traffic is exchanged by ISPs to support user flows that traverse multiple domains. In this case each ISP is in charge of a network domain. Pricing for used resources may use models like peer-to-peer, wholesale or retail. Schemes like edge pricing, where an access ISP pays only to the higher level ISP to relay traffic have been proposed. Such pricing models utilise long-term SLAs, which avoid per-call signalling overhead. On the other hand distributed pricing involves negotiating the tariff for a call or a flow on a hop-by-hop basis with each involved domain. The signalling overhead is higher for this scheme, and may cause scalability problems in the core of the network [24]. Aggregation of calls may be used to improve scalability in the core network.

There is a close relationship between inter-domain pricing and routing. Routing mechanisms may attempt to find an end-to-end path that minimises service cost, or a path that meets QoS requirements for an application. Network domains require resource brokers to negotiate with peering domains to establish routes that meet the QoS and price constraints.

Resource usage accounting is required for charging and billing. Accounting can be performed for the network transport service as well as content [24]. Various schemes can be used for transport service accounting depending on whether an intra- or inter-domain scenario is being considered. These include recursive accounting, where payments made by users connecting through access domains are propagated to upper level network providers in the chain towards the corresponding node. This scheme can support both sender and receiver payments.

2.5.2 Pricing Profiles

The use of profiles in pricing of communication services aims at enabling users to influence the quality-to-cost relationship of services rendered to them. Generally pricing profiles are tied to service classes available on the target network. Moreover, service classes can

be associated with different access networks. Thus a best effort service class can be mapped to a lower pricing profile on a network where QoS guarantees are offered. The same service class can be mapped to a WLAN where QoS guarantees are not offered. The service classes specify the level of QoS guarantee using service level specifications (SLS). Each service class differs in predictability, reliability and efficiency [42].

In related work [14], the use of profiles in pricing of 3G services was explained. The paper introduces a silver flat-rate priced profile suited for best effort transport, a gold fixed tariff rate profile suited for lightly loaded networks with guarantees on a minimum level of bandwidth, and a dynamically priced platinum profile that offers hard guarantees on bandwidth subject to increasing tariffs during congestion periods. Connection admission control is used to prevent excessive congestion in classes used for the platinum and gold profiles. Admission to those classes is bound to the availability of usable resources allocated in a DiffServ style from a pool of sharable resources. Through inter-class resource sharing, unused resources from one class can be allocated to another to enable admission of more users or support improvement of performance levels. Refer to [14] for more information on these pricing profiles.

2.5.3 Demand and Supply of Network Resources

The supply of network resources is meant to meet a certain level of demand for resources. Estimations of average demand depend on resource usage measurements. Bandwidth is the main metric that characterises the available resources in a network. In wireless networks the available bandwidth is affected by varying channel conditions and transmit powers of nodes in a given cell and neighbouring cells. Transmit power influences interference levels and affects channel data rates [43]. The theoretical maximum rate at which error free bits can be transmitted over a bandwidth limited channel in the presence of noise is given by Shannon's law as: $C = W \log_2(1 + \frac{S}{N})$. C is the channel capacity in bits per second, W is the bandwidth in hertz and $\frac{S}{N}$ is the signal-to-noise ratio [44]. The demand for bandwidth in communication networks differs in access networks and core networks. Demand is greater at network edges than in the backbone. This is attributed to the localisation of most traffic to access networks.

In wireless communications access networks form the bottleneck on the end-to-end link between the communicating nodes. This is more as a result of the different technologies

used for the air interface between user equipments and the base station, and the predominantly fibre optic fixed line backbone. To meet the increasing demand for bandwidth by mobile applications developers of mobile broadband systems have identified two key technologies with the potential to meet operator needs in the near future. The 3GPP long term evolution (LTE) and WiMax can be integrated or deployed independently to provide greater amounts of bandwidth as compared to present networks. According to forecasts the demand for mobile bandwidth will come mainly from multimedia applications like Video on Demand (VoD).

Globally mobile data is expected to double every year through 2013 [45]. Mobile handsets supporting speeds higher than typical 3G handsets, i.e. 3.5G and portable devices like laptop aircards will account for more than 80% of this traffic. Almost 64% of the mobile traffic will be video. There are a number of methodologies used in estimating the growth in demand for network bandwidth. The service adoption model, for example forecasts the edge demand based on the adoption of services and statistical information like population growth and previous adoption rates for comparable services [44]. On a long time frame bandwidth demand (offered load at network edges) is expected to grow steadily. On a short time scale the average peak demand across a 24 hour day starts rising rapidly at the start of business hours; it peaks at about 10 am and will drop slightly at the end of business hours. A spike in demand can occur around 8 pm [44]. This can be attributed to the use of digital home entertainment systems.

Sufficient network capacity (supply) is required to meet the demand for bandwidth. For circuit switched networks the capacity is determined by the number of voice calls that can be supported. Capacity in packet switched networks is a measure of the peak load the network can handle. This definition holds in classical best effort packet switched systems, where more users and user sessions can be added at the expense of network performance. A limit to the number of users, specifically the number of guaranteed performance sessions can be set to ensure a certain degree of performance. In convergent network employing a combination of both circuit and packet switching the capacity of the network is determined by the intelligence of the systems to differentiate services and manage traffic [44].

Network capacities are required to increase 8-fold in order to meet doubling demand for network resources [44]. Thus, a substantial capital investment is required to provide extra capacity. Operators are normally motivated by revenue potential in investing in

the supply of extra bandwidth. These investments are affected by technology, competition, financial and regulatory factors. Operators need new infrastructure to support convergent services, new and existing income generating services, minimise service implementation costs, while adhering to regulatory and industry standards.

Revenue generation motivates new and incumbent operators to supply bandwidth. Bandwidth supply can be done to maintain high marginal revenue, and/or generate new revenue streams to offset declining revenue streams. Incumbent operators can increase their network capacity to maintain revenue streams, especially when new entrants create new capacity to lure customers. However, the turn-around period on new investments is an important factor; thus, operators will save on capital if they can create new services on existing infrastructure. Noting that revenue growth rates can fall below CAPEX growth rates operators continuously need ways to increase network capacities and sustain short term income flows without investing heavily in capacity expansion. Service outsourcing presents a solution to short-term capacity expansion needs to meet surging periodic demand.

Utility Function

Utility as a measure of the value users derive from services rendered to them. Users are affected by budget constraints, and will adapt their usage patterns to changing network prices in order to maximise their utility, subject to minimum quality requirements. The utility of a network service can be viewed in terms of a set of transmission parameters [25]. The transmission bandwidth and QoS parameters directly affect it; thus they can be adapted for its maximisation. With a high utility users get a big value for their payments, since the surplus of the utility versus the service cost is high. Utility functions follow a model of diminishing returns, i.e., the marginal utility as a function of bandwidth diminishes with increasing bandwidth [25].

2.6 Service Provision Architectures

Service provision architectures constitute the functions and building blocks for service management and control [46]. Service outsourcing requires a service provision platform with well defined call or session control procedures, and involving mobility and handovers

between wireless networks. The IP Multimedia Subsystem as standardised by the 3GPP is used as the service provision architecture in this thesis. Various protocols have been standardised to facilitate IMS operations to achieve AAA functions, mobility and provide required QoS levels for sustaining acceptable QoE for users.

2.6.1 Service Control

Service delivery involves end-users, network providers and service providers. Information input is needed from each of these parties in order to facilitate service provision. Service control comprises of service control points and service execution points. The network provider (or network operator) occupies a critical position with regard to the two functions. Interaction between the control and execution points of service control occurs in a request-response format. The response time is critical to system performance. When speedy response is required, real-time user input may not suffice; thus, it is necessary to involve an agent that would act on behalf of the user in the network provider's domain. Inter-network service control agents are also required.

2.6.2 Service Phases

Network service control can generally be categorised in three phases: the setup, execution and termination phases. Service outsourcing transactions should occur within the three phases.

Set-up phase

In service delivery scenarios where a user receives services from a network, the user would present to the network provider the relationship between his needs and his "willingness to pay" for requested services [46]. Networks and users normally agree on the quality and cost of service before commencing service delivery. Similar agreements are made in network-to-network interactions. Discovery of available services on a network is part of the setup stage. Generally service discovery precedes cost negotiations. It is after service discovery that service level specifications can be made followed by formulation of SLAs. The consumer's view on the quality-to-cost relationship represents the subjective value of the service. The provider presents its own pricing policy in terms of tariffs and service

quality. Service delivery and charging can only be triggered if the views of the consumer and provider are identical. This is often hard to achieve; hence, a resolution mechanism is required. The mechanism may involve renegotiation procedures by adjusting the terms presented by the parties. The drawback to this for user to network transactions is the requirement of intelligent mechanisms at the consumer and provider domains [46].

In network-to-network interactions the deployment of intelligent agents to negotiate service delivery parameters for cost-to-quality relations is highly possible. The additional signalling overhead should, however, be investigated. In network sharing scenarios the service cost acceptance terms for the home operator are limited by the charge to be passed on to the final user. Thus, the terms defined in the policies of the visited network should tally with this major requirement. Alternatively, the opinion of the network provider (visited operator) can be accepted 'as is', and that of the consumer (home operator) disregarded, or taken as a hint in determining the provider's policy. This solution may be simple to implement, but it is not flexible for the home operator [46].

Service Execution Phase

Operators will control service delivery by using management mechanisms. During service delivery charging information can be obtained from the network at specific intervals or upon occurrence of specific events. In networks where admission control is enforced to prevent and control congestion, monitoring of the load level will occur in the execution phase. This will allow the network to dynamically control the cost-to-quality relationship in the survive delivery agreement. The decision to increase the capacity of the network depends on analysis of information gathered from transactions and events that occur in this phase.

Service Termination Phase

Service termination occurs at the end of the service session. Reports on network usage by the session are sent to network entities in charge of processing for analysis, charging and billing purposes. This information can be used to verify that the service was delivered to the end-user according to the agreement. In network-to-network transactions, this information can be used to assess adherence to SLAs governing service delivery. In

case of service failure the contractual agreement between the networks may include compensation options [46].

2.7 Discussion

There is potential for network operators to reduce CAPEX and OPEX related to provisioning of network resources through network sharing. A number of network sharing solutions have been deployed in commercial networks. Network sharing solutions are guided by SLAs between operators. The SLAs spell out parameters for service quality, i.e. SLS, cost of service, geographical scope and duration of the agreement. Long term SLAs are characteristic of network sharing solutions; hence there is lack of flexibility and adaptability. The performance of network applications depends on current network conditions, which depend on the supply and demand of network resources. The supply of network resources is achieved through network dimensioning done in the planning stage. Generally networks are dimensioned to meet the highest value of average demand for resources measured over different periods. Demand for network resources is highly dynamic and can exceed the available resources. This is common in highly competitive locations, where most operators have coverage.

Whereas most network sharing solutions aim at solving the lack of network coverage, lack of sufficient capacity during peak periods and when some access points are out of service is still not addressed. Network churn is the main product of service dissatisfaction due to failure of congested networks to meet service requests. A network sharing solution tailored to respond to changing network conditions, e.g., congestion levels, and designed to achieve virtual capacity expansion and minimise blocking of service requests or quality degradation is proposed and developed in this thesis. Critical financial conditions are required to ensure the feasibility of such a scheme. In general each network tries to balance the supply and demand of network resources by provisioning and selling to users and other networks. Moreover, service control mechanisms that support real-time, per-session negotiation of SLAs between networks participating in resource sharing are required.

The IMS standardised by the 3GPP has been chosen in this thesis as a service provisioning architecture for developing and evaluating service outsourcing. A review of important

information on IMS is given in the next chapter. Emphasis is given to charging control, signalling transactions related to mobility, charging and inter-domain interactions.

University of Cape Town

Chapter 3

Background of Charging Control in IMS

The IMS was introduced in release 5 of 3GPP standards as a platform for control and delivering multimedia services over IP network [5, 47]. Moreover, in release 8 of 3GPP specifications the evolved packet system (EPS) was standardised. This comprises a Long Term Evolution (LTE) access network, a non-3GPP access network and an evolved packet core [30, 48]. This versatile and robust network architecture is designed to promote rapid creation and deployment of services in the paradigm of services architecture evolution. Charging of these services is required to enable operators and 3rd party providers to recover costs related to service delivery. The 3GPP has standardised charging control for IMS and EPS [10, 49]. Prior to release 7 of 3GPP standards, policy and charging control were handled separately as Service Based Local Policy (SBLP) and Flow Based Charging (FBC) [50]. A unified Policy and Charging Control (PCC) architecture was adopted for IMS in release 7 and later for EPS in release 8 [50, 51]. The Online Charging System (OCS) and the Offline Charging System (OFCS) have been standardised by 3GPP.

Research and development work on IMS and EPS has been going on for a number years and is reaching a maturity stage. The Open IMS core developed by the Fraunhofer institute in Germany is a good reference for outcomes of IMS research. The UCT IMS client and IPTV framework are important developments in the research chain. These standards provide a foundation for research on IMS charging and outsourcing. However, the lack of a standards compliant IMS charging framework to complement these core elements is evident.

PCC introduces two important entities; the Policy and Charging Rules Function (PCRF) and the Policy and Charging Enforcement Functions (PCEF). Moreover, new interfaces and protocols are introduced and adopted. Diameter [7] is the main protocol standardised for information exchange between different PCC entities. Charging in 3GPP networks is accomplished at the transport and service planes. PCC entities are located at various planes to facilitate charging control, which can be mapped to duration of service usage, traffic volume or other charging metrics.

An Application Function (AF) and a Subscription Profile Repository (SPR) are additional entities used in charging control. The AF interacts with applications to extract session information and send it to the PCRF across the Rx reference point [52]. The SPR stores user subscription information that is used in the creation of PCC rules by the PCRF. Events occurring on the transport plane can be sent to the PCRF by the PCEF, and forwarded to the AF. Figure 3.1 illustrates the architecture for policy and charging control in 3GPP networks [53]. The different reference points used in policy and charging control are shown.

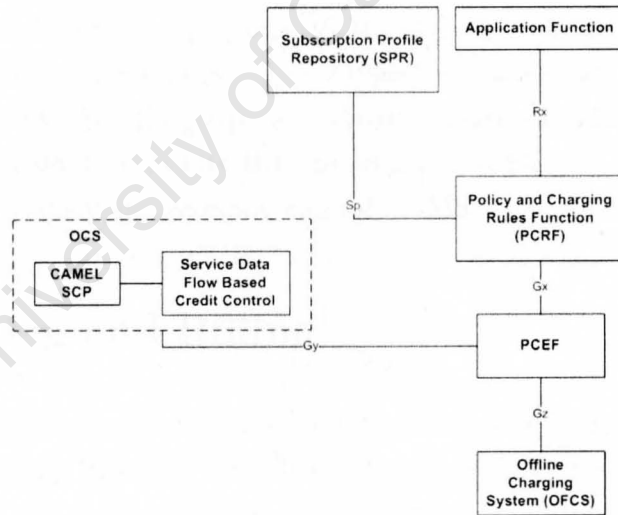


Figure 3.1: Policy and charging control architecture

The PCRF is the central point for policy and charging control; it uses information received from AF, PCEF and SPR to make policy decisions to be enforced by PCEF. In addition to enforcing policy decisions the PCEF conveys access specific information to the PCRF across the Gx interface. For traffic plane charging the PCEF interacts with the OCS for credit control and OFCS to report resource usage.

PCC involves gating control, QoS control and charging control. Gating control involves allowing or blocking of IP packets belonging to a flow based on network events, e.g., events occurring at the AF. QoS control involves enforcement of PCRF policies, e.g., authorised QoS at the PCEF. Charging control on the other hand involves online credit management by the OCS, which assigns credits for particular IP flows to influence access control [50]. For more information on 3GPP policy and charging control, please refer to appendix B.2.3, which provides supplementary information on PCC procedures.

This chapter reviews charging procedures for IMS and EPS and provides details of protocols used for charging control as well as protocols used in IMS session control. QoS policy control is not discussed since it has been exhaustively addressed in other research work [54]. The information presented in this chapter forms a foundation for the design of a charging and outsourcing framework implemented in this thesis.

SIP [6] is a text based protocol used for session control in IP communications. It is used for communication between user agents and proxies by conveying session control information, which may be carried in SIP headers or as special bodies encoded in specific formats like Session Description Protocol (SDP) [47]. Its application to IMS communications is paramount, hence it has been utilised in the design of the charging and outsourcing framework. For the purpose of clarity, additional information on IMS session control is provided in appendix B.1. In addition to SIP the Diameter protocol is used extensively for policy and charging control in IMS.

3.1 Charging and Billing

Charging is the process of collecting information on resource usage and forwarding this information to a billing function. In 3GPP charging is defined as a function whereby information related to a chargeable event is formatted and transferred to a billing function for use in determining user bills [55]. It requires service rating to determine the cost of the service. Billing is the process by which resource usage is associated with the user to generate bills.

With respect to charging end users, network operators and third party service providers have different requirements. End users prefer receiving one itemised bill for all services accessed through the same network. Moreover, they prefer a simple and consistent

charging methodology. Operators require that the charging scheme covers all aspects of services rendered on the network to ensure that all related costs are recovered. Moreover, the charging scheme should be efficient and scalable [56]. Scalability is a very important factor in accounting procedures. More accurate representation of resource usage by a flow requires granular collection of information about it.

Granularity and scalability are opposing aspects, i.e., a higher granularity leads to a less scalable system. A realistic system is achievable by pushing granularity towards the access networks, and scalability towards the core network [24]. This is because access networks carry less traffic than core networks, thus a lot of processing can be performed on traffic in access networks without impacting on service delivery. Third parties require transparent and automated apportioning of revenue for content and services provided over the network. A comprehensive charging system requires control entities at different planes of the architecture to meet these requirements. Moreover, interaction between the charging entities using standardised protocols is needed.

Service outsourcing relies on charging mechanisms due to the proportions of revenue to be shared between custodian and candidate networks. Outsourcing can be used with IMS online and offline charging enforced at the transport plane and service planes. The transport plane is divided into an access network domain and the core network domain. The access domain comprises various IP Connectivity Access Networks (IP-CANs) like WLAN and E-UTRAN [53].

3.1.1 Diameter Protocol

Diameter is emerging to be the main protocol for Authentication Authorisation and Accounting (AAA) for Internet communications. Diameter has been standardised by IETF [7] and specified as a base protocol to be implemented in all its applications, which meet specific requirements e.g., Diameter credit control application [57] for online charging. Each Diameter application is assigned a unique application ID. The base protocol is extended by addition of new commands or Attribute Value Pairs (AVP), which are identified by command codes per Request/Answer pair.

The Diameter base protocol provides the function of delivering data in AVPs, which can be added arbitrarily to transport any type of information between Diameter nodes. They contain an AVP code, flags, a vendor ID etc. Service outsourcing in IMS relies

on Diameter to exchange offer price, outsource duration and other information between custodian and candidate networks.

The Diameter architectural structure includes clients and server instantiated as proxies, redirects and relays. It supports roaming and inter-domain AAA services. Different domains are differentiated as home or foreign realms. In Diameter messages the realm is the part of the URI that comes after the '@' sign [7]. Routing of Diameter messages is based on Destination-Realm and Destination-Host AVPs.

3.1.2 Diameter Credit Control Application

The Diameter credit control application can be used for real-time interaction between an application function and a credit control server [57]. It involves verifying the availability of credit, reservation, deduction of credit at service completion, and refunding of unused credit. The 3GPP OCS is an example of an entity that functions as a Diameter credit control server.

The AF takes the role of a credit control client that receives credits from the server and monitors the usage. It reports the used credit quota by sending interim reports and can request new credits during the life of a session.

The credit control application uses two message commands, i.e. Credit-Control-Request (CCR) and Credit-Control-Answer (CCA). Diameter credit control is achieved either by credit reservation or by direct debiting on the user's account. The allocated credits translate to an allowed data volume, duration of service usage, or service units, e.g. a number of messages. Simultaneous credit authorisation requests can be sent to the credit control server. In this thesis Diameter credit control is used in implementing the Ro reference point for the online charging system.

3.1.3 Flow Level (volume) Charging

Flat rate and duration charging are popular for voice telephony services offered on telecommunications networks as well as the Internet. Volume charging requires the collection of accounting information for service flows traversing the network [51]. The charging system developed in this thesis supports volume charging at the media plane,

where the PCEF resides. It is used in service outsourcing to account for bandwidth usage on candidate networks during the outsource period.

3.1.4 Content Charging

Content based charging is required for applications delivered on the network. IPTV services in IMS are an example of these services. The charging and outsourcing framework is evaluated using IPTV services, thus providing a practical platform for research into content charging.

3.2 IP Multimedia Subsystem

The 3GPP IP Multimedia Subsystem (IMS) is a service convergence framework that provides a cost-effective and network controlled platform for the design and delivery of IP-based multi-media services [58]. IMS communications utilise existing IP networks and protocols standardised by the IETF. IMS depicts a horizontally layered architecture that comprises an application (service) plane, a control (signalling) plane and a transport (media) plane. The application plane hosts different application servers e.g., presence, IPTV and conferencing.

The main components of the control plane are the Call Session Control Functions (CSCF), which perform registration of users and the establishment and control of IMS sessions. There is a home subscriber server (HSS) used to store user and network related information e.g., user profiles. Other entities belonging to the control plane include gateways that perform the translation of signalling information between different network architectures. The transport plane contains network nodes that lie in the path of media traffic between the communicating equipment e.g. two or more user agents (UA) or a UA and an application server. Figure 3.2 illustrates a simplified view of the IMS architecture [5]; only entities that are used for outsourcing and charging processes in this thesis are discussed below.

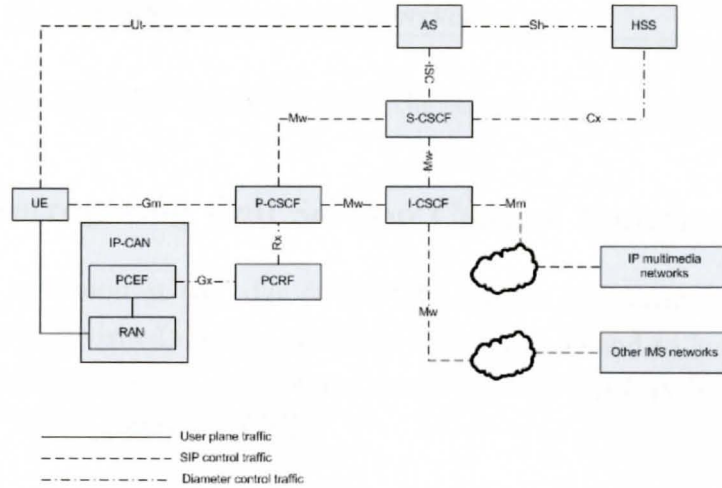


Figure 3.2: Basic entities of the IMS architecture

3.2.1 Proxy Call Session Control Function (P-CSCF)

The P-CSCF is the first point of entry to the IMS network. It is traversed by all requests generated by the user equipment and requests from the network to the user equipment [47]. For the purpose of policy and charging control, the P-CSCF can extract session information from SIP signalling and send it to the PCRF over the Rx reference point. In this regard it acts as an Application Function (AF). The serving P-CSCF is allocated during IMS registration. This fact has an important on service outsourcing transactions, since an re-registration would be needed to accomplish handover to the candidate network.

3.2.2 Serving Call Session Control Function (S-CSCF)

The S-CSCF is a central component in the signalling plane and acts as the registrar entity of the IMS. It maintains a binding between the IP address of the user's terminal and the Public User Identity (IMPU). The S-CSCF downloads user profiles and authentication information from the HSS at registration time [5]. During session establishment it intercepts all SIP signalling from the client and uses the profile information to determine, based on the filter criteria, if the signalling should be forwarded to an application server. Services like IPTV rely on the HSS to forward SIP requests to the IPTV AS. For example if a SIP INVITE contains the address of the IPTV AS in the To header, the

messages will be routed to the IPTV AS if the service point trigger is matched.

3.2.3 Interrogating Call Session Control Function (I-CSCF)

The I-CSCF is located at the edge of the IMS domain [47] forming an entry point of SIP signalling in a visited domain. Its main function is to find and locate and assign an S-CSCF for IMS registration. Moreover it can be used for topology hiding purposes (i.e., as a topology hiding gateway - THIG).

3.2.4 Home Subscriber Server (HSS)

The HSS is an evolution of the Home Location Register (HLR); it is the central repository for user related data [47] that includes user identities, registration information, access parameters and service triggering information [5]. The HSS is used to define application servers and initial filter criteria for routing SIP messages to respective application servers. The initial filter criteria require service point triggers that define matching SIP headers of messages to be routed to an application server. A sample configuration of the HSS service point triggers and initial filter criteria is given in appendix E.1.2. Charging and service outsourcing mechanisms rely on HSS functionality to facilitate IPTV services. Furthermore, charging for VoIP services uses an independent charging server that uses the HSS to configure routing of SIP messages.

3.2.5 Home and Visited Domains

With respect to roaming IMS network access adopts mechanisms of 3G communications. Most IMS call control entities are located in the home network; however, the P-CSCF can be located in the home or visited network [47]. IMS users connect through an IP-CAN, which can be provided by the home or visited network operator. In the 3G architecture the location of the P-CSCF would be relative to the location of the GGSN, which could be in the home or visited network [59]. Service outsourcing exploits IMS roaming architecture to position the location of custodian and candidate P-CSCFs.

3.2.6 IMS Charging Control

IMS charging standards are created by the 3GPP with most of them being extensions of IETF standards. 3GPP has defined a common charging framework for all IMS services and applications [49]. There are additional standards tailored for online and offline charging at different IMS planes. In online charging there is a real-time interaction between charging control mechanisms and resource consumption. On the other hand, in offline charging there is no real-time interaction between charging control and resource consumption. With respect to billing, pre-paid and post-paid options are supported by network operators.

An IMS charging architecture is depicted in Fig. 3.3. The OCS, OFCS and reference points for online and offline charging are shown.

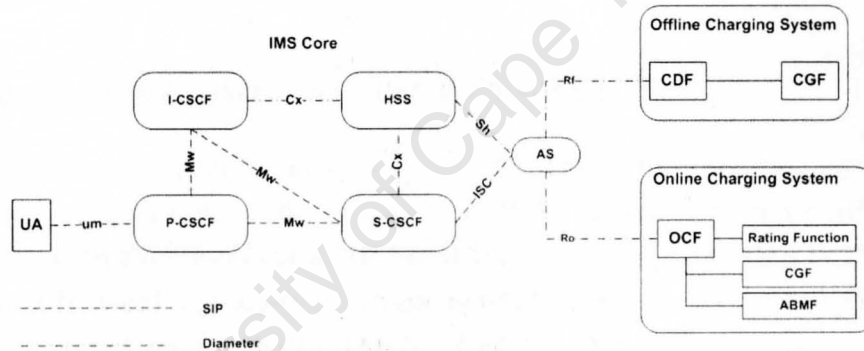


Figure 3.3: Service level charging in IMS

The charging system requires a rating (pricing) function that provides information on the cost of services and content. Tariff information can be stored in a price look-up table and contains tariffs for various pricing plans e.g., economy, standard or premium. The pricing plan applicable to every user would be stored in a customer profile table in the billing domain. This table would also contain the discount types, customer status and user ID [10]. Charging detail records (CDRs) containing the user ID, time and duration of service delivery, type of service or content, access network information, transaction type and volume of data transferred are generated during a charging session or event.

For flow level charging, IMS charging control is performed alongside policy control. In this case the P-CSCF or an AF extracts session information from IMS signalling and sends it to the PCRF. The extracted information includes: IP addresses and other

information used to identify the flow. QoS information like bandwidth, media type and media codecs is also extracted. The PCRF uses this information in addition to subscriber profile information to formulate PCC rules that are installed and enforced, in form of IP packet filters, at the PCEF. Figure 3.5 illustrates the creation and use of PCC rules in IMS [50]. The session information is extracted for SIP signalling at the P-CSCF and sent to the PCRF for PCC rule creation. Once installed at the PCEF, which resides in the Packet Data Network Gateway (PDN) GW, the rules are enforced for gating, QoS and charging control. With respect to charging, the PCC rule includes a charging key that is used to determine the applicable tariff, the charging method i.e. online, offline or no charging, and measurement method e.g. volume or duration. More information on charging in IMS, including various reference points is given in appendix B.2. A subset of these charging transaction are used in designing and implementing the IMS charging framework.

Charging for IMS Applications - IPTV Charging

IPTV is a service that can be integrated into the IMS by incorporating an application server to control and deliver the service. The IPTV application server (IPTV AS) is made available to IMS users via a SIP based ISC reference point. IMS IPTV has been standardised by various groups; thus, various IPTV architectures exist. IPTV benefits from the enhanced service control capability of the IMS, e.g. QoS control, charging and billing [60]. Figure 3.4 illustrates a high level view of a TISPAN IMS-based IPTV architecture [60].

The IPTV application server provides service information to the UE, while the Media Control Function (MCF) selects and controls content distribution to the Media Delivery Function (MDF). The MDFs are responsible for the delivery of media to the UE. There are three types of IPTV services, i.e., broadcast services, unicast services e.g., VoD and Personal Video Recorder (PVR) [60]. The IPTV scenario explored in this thesis implements unicast (VoD) services.

Charging for IPTV services involves real-time interaction between a Charging Trigger Function (CTF) that is integrated into the IPTV AS to capture defined charging triggering events e.g., an INVITE requesting of a media channel and charging system entity. This occurs across the Ro and Rf interfaces for IMS online and offline charging respectively. Standardised event based charging mechanisms [61] can be suitable for video on

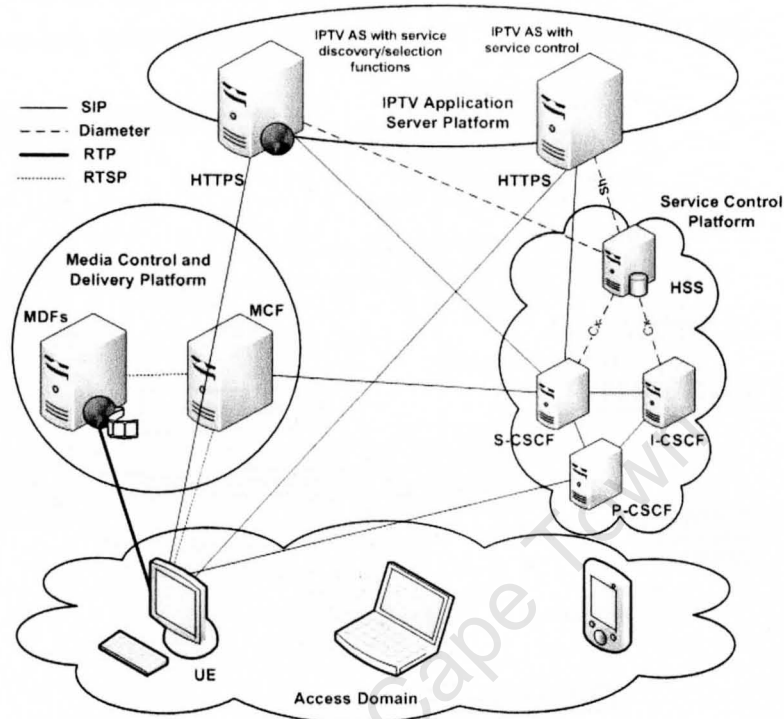


Figure 3.4: TISPAN IMS IPTV architecture

demand charging. In addition to content charging, duration, volume and QoS charging are appropriate for IPTV services.

3.3 Evolved Packet System

In EPS the Evolved Packet Core (EPC) presents a single core network for multiple heterogeneous access networks. The EPS architecture requires mobile IP in the EPC to support mobility between the heterogeneous access networks. The EPC comprises the following entities: a Mobility Management Entity (MME), a Serving Gateway (S-GW), a PDN GW, an Access Gateway (A-GW) and an Evolved Packet Data Gateway (ePDG). Other entities include the 3GPP AAA server, Border Gateway (BG), HSS, PCRF and SGSN [53, 62]. The MME is a control plane entity within EPS supporting authentication, roaming, PDN GW and S-GW selection etc. It also supports inter-core network node signalling for mobility between 3GPP access networks, and bearer management. The S-GW interconnects the E-UTRAN and the EPC. Among other functions it acts as a layer

2 or local mobility anchor point for inter-eNodeB handover and inter-3GPP mobility. PDN GW provides an interface for 3GPP and non-3GPP access networks towards the operator's IP services network, e.g. IMS across the SGi interface [62, 63].

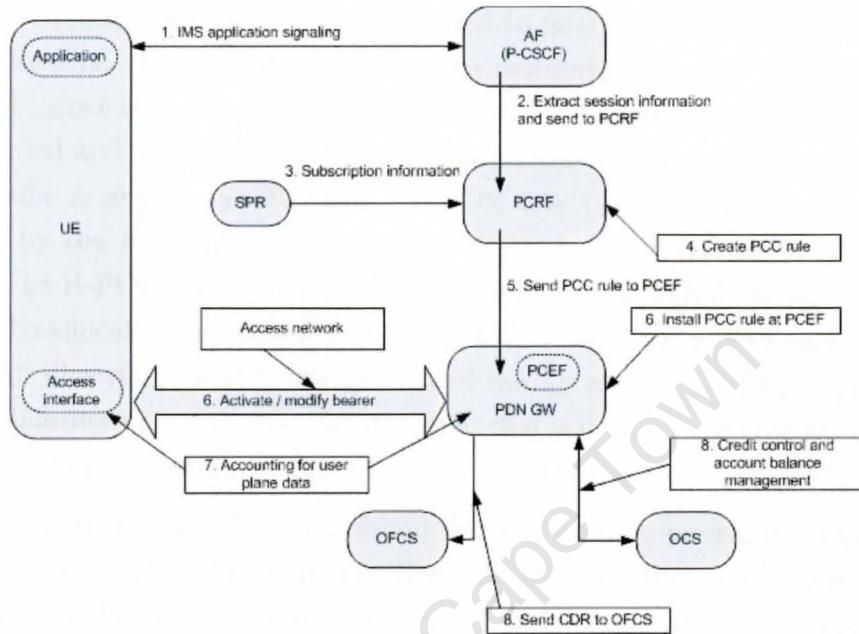


Figure 3.5: Creation of PCC rules in EPS

In EPS the SGSN performs inter-EPC node signalling to support mobility between 2G/3G and E-UTRAN access networks. It also performs PDN and S-GW, and MME selection for handovers to E-UTRAN. Access to EPC by non-3GPP access networks is provided by A-GW and ePDG. Non-3GPP access networks can be trusted or untrusted. The A-GW is used by trusted non-3GPP networks, while ePDG serves untrusted networks [62]. 3GPP AAA server resides in the H-PLMN and serves non-3GPP access network users. It facilitates AAA functions and co-ordinates information needed for mobility between 3GPP and non-3GPP networks. A 3GPP AAA proxy resides in the VPLMN to perform similar functions. This server selects a S-GW for the UE during the attach process.

EPS can benefit from service outsourcing by leveraging the support for mobility among different access networks with varying levels of bandwidth.

3.3.1 Roaming in EPS

The S9 reference point has been standardised to interconnect PCRFs in the home and visited domains. Control of services during roaming is handled by the home PCRF (H-PCRF) since the visited PCRF (V-PCRF) cannot modify session information. Both home routed and local breakout access scenarios may be used. In home routed all user plane traffic is sent to the PCEF in the PDN GW in the home network. PCC rules are sent by the H-PCRF to the visited BBERF in the S-GW or A-GW via the V-PCRF. The H-PCRF also enforces the PCC rules at the PCEF. When local breakout is used dynamically provisioned PCC rules will be enforced at the visited PDN GW. Figure 3.6 illustrates various interfaces used during roaming in EPS. Appendix B.2.6 provides information on the S9 reference point that is used between the home and visited domains to facilitate charging control during roaming.

Service outsourcing in IMS uses a subset of EPS roaming mechanism that essentially are an enhancement of IMS roaming. However, only a subset of EPS entities that are applicable to IMS communications are utilised in designing the outsourcing framework.

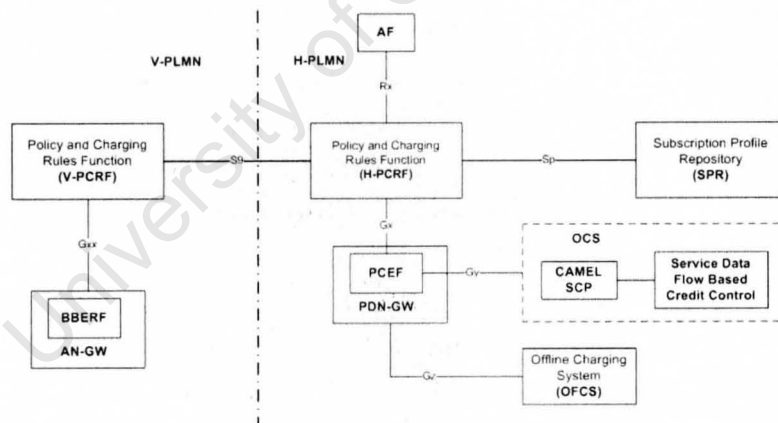


Figure 3.6: S9 policy and charging control reference architecture

3.4 Conclusions

The IMS provides a platform for delivering a range of multimedia services. IMS research has been going on alongside the evolution of mobile networks from 3G (UMTS) towards

4G (EPS). The IMS comprises a media plane, a control plane and an application plane inter-connected with standardised interfaces, with each plane providing a set of services to the other. IMS communications rely on SIP and Diameter as core protocols. SIP is used to transport control information between IMS clients and proxies, while Diameter is used for policy and charging control, and for various interfaces between IMS entities. Charging is standardised for the IMS; the charging mechanisms have evolved from 3G charging mechanisms. Charging is supported for all planes of the IMS. For example volume or flow level charging is achievable at the media plane. For media transport IMS relies on the real time protocol and the real-time transmission control protocol.

Service outsourcing exploits IMS control mechanisms and the the inter-operability between different networks that use IMS for multimedia service control. For instance an IMS session can be outsourced from a U-TRAN network to an E-UTRAN access network to gain from more bandwidth.

Chapter 4

Network Capacity Expansion Through Service Outsourcing

Service outsourcing is a resource sharing scheme designed to enable operators to request or lease network resources in real-time. As mentioned earlier, the main motivation for service outsourcing is enabling operators to accept and service requests from subscribers during periods when the demand for resources exceeds the network resource capacity. In this chapter the design of a service outsourcing framework for wireless networks is modeled and presented. The chapter starts by addressing the resource pricing conditions and challenges surrounding service outsourcing. This is followed by the architectural and technical aspects.

4.1 Service Outsourcing Concepts

Service outsourcing involves two or more operators who engage in resource sharing agreements. Unlike conventional resource sharing agreements that are defined by long term SLAs, service outsourcing SLAs are negotiated and executed in real-time for each session. This occurs when the custodian operator presents a request for resources to a candidate network.

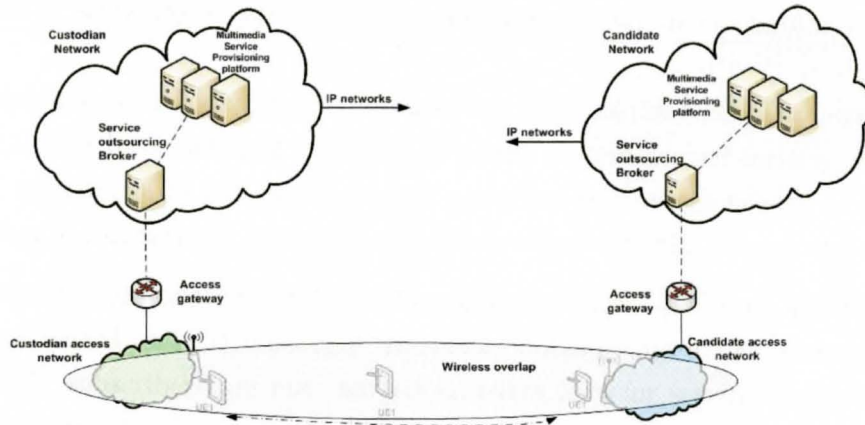


Figure 4.1: Service outsourcing general architecture

4.1.1 Service Outsourcing Scenario

The number of subscriber sessions that can be admitted to a wireless network is limited by the amount of bandwidth available in the air interface, as well as backhaul links and other links on the path between corresponding subscriber equipment. Blocking of resource requests occurs when attempts to establish sessions are made on a congested network. A network is considered to be congested when the level of resource utilisation has reached a threshold above which it is impossible to meet quality guarantees of active sessions. For session based services such as multimedia calls, congestion can be associated with the active number of flows. Generally the throughput of a network gives a measure of the congestion level. Moreover, the efficiency of network operation gets compromised. Instead of rejecting or holding new resource requests, service outsourcing enables the congested network to lease resources from a reachable network. An outsourcing request is sent to the candidate network to establish an agreement for the transaction. The custodian network offers to pay a specific fee to lease a certain amount of resources over a period not exceeding a fixed length of time.

4.1.2 Requirements and Conditions

Service outsourcing is governed by technical and fiscal conditions. The main technical requirement is the existence of a wireless overlap between the custodian and candidate access networks at the location of the user equipment. This is depicted in Fig. 4.1, where

the custodian and candidate access networks are reachable by UE1. The candidate access network should have sufficient resources to admit the outsourced service. A mechanism for joint call control by the candidate and custodian networks is needed to facilitate the monitoring of call activity to ensure each party adheres to the outsourcing SLA.

Fiscal requirements relate to profitability goals of network operators, and subscriber expectations regarding service charges. In service outsourcing, unlike conventional roaming [35, 37], subscribers are not charged an extra levy for services accessed using the candidate network. The use of a constant predictable service fee requires sharing of revenue from subscriber payments.

We note that service delivery by Mobile Virtual Network Operators (MVNO) sees users being charged a predictable fixed fee [35]. However, in the MVNO scheme network-to-network charges are negotiated on long term basis, whereas service outsourcing involves real-time cost negotiations. Traditionally subscribers are billed by the custodian (or home) operator. Apportioning of revenue between both operators is performed by the custodian operator.

4.1.3 Service Outsourcing SLA

As mentioned earlier, a service outsourcing agreement is formulated for every session prior to handover to the candidate network. The main parameters of the SLA are the offer price (OP), outsource duration (T) and the required QoS resources. In this thesis, bandwidth is the main QoS resource leased on the candidate network. The SLA parameters are sent by the custodian network in the outsource request. The OP indicates the fee to be paid by the custodian network for resources used by the session. The OP can be fixed for outsourcing transactions between the operators. In this thesis we explore a scenario where the OP is negotiated for each session, while taking into account changing network conditions in the pricing of leased resources.

The outsource duration specifies the expected lease duration for sessions. The effective lease duration (t) is affected by user actions, i.e., users could terminate the session or move out of the coverage range of the candidate access network. In that case the lease duration would be less than the initial negotiated value T, i.e. $t < T$. Bandwidth requirements should be met on the candidate network for the duration of the outsource period. By admitting the outsourced session, the candidate network commits to provide

the leased resources. Penalties may be enforced if either operator fails to adhere to the outsource SLA.

The scope of the SLA can cover all cells of the candidate network, or can be confined to the current cell. This implies that in case the user moved out of the coverage range of the candidate access network, session continuity would be guaranteed if specified in the SLA. There would be no guarantee for session continuity if the SLA only covered the cell of entry to the candidate network.

4.1.4 Service Outsourcing Transactions

Service outsourcing involves a setup phase, an execution phase and a termination phase. SLA negotiation occurs in the setup phase, while service delivery occurs in the execution phase. SLA renegotiation would occur in the execution phase if lease renewal were to be done at the end of the outsource duration. Service outsourcing enters the termination phase when the session is terminated. There is a pre-setup phase where the UE detects reachable access networks and sends information to the custodian network.

Service Outsourcing Setup Phase

Within the set-up phase the custodian network and the candidate network come to an agreement on the service quality and the associated charge for service delivery, i.e., the quality-to-cost relationship. In this respect the custodian network takes the role of the consumer and the candidate network assumes the role of the provider.

After receiving a resource request from a subscriber the custodian network initiates service outsourcing if, due to congestion, the session cannot be admitted to the network. The candidate network is selected from a computed list of reachable access networks that had been reported by the subscriber's UE. Thereafter a service outsourcing request is sent to the candidate network.

On receiving the outsourcing request the candidate network can accept or reject it based on the value of the OP. If the OP is too low a challenge to the outsource request may be sent to the custodian network, which would determine the suitability of the counter price before updating the OP and sending a final outsource request for the session.

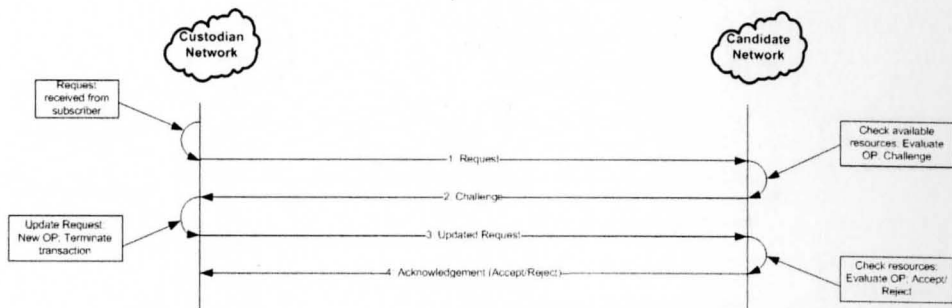


Figure 4.2: Service outsourcing - setup phase

When the second outsource request is received the candidate network either accepts or rejects it based on the new OP. The first request would have been accepted if the OP were acceptable to the candidate network. On receiving the challenge, the custodian network could terminate service outsourcing if the counter OP were too high and unacceptable. An acknowledgement is sent by the candidate network if the request is accepted.

Figure 4.2 illustrates the three step service outsourcing session setup phase. The custodian network initiates service outsourcing (message 1); after evaluating the OP and checking for requested resources, the candidate network sends a challenge (message 2). The custodian network updates and sends a new request (message 3); the update could terminate the outsourcing transaction. After processing the updated request the candidate network sends an acknowledgement, i.e., rejecting or accepting the request (message 4).

Session establishment information to support AAA functions for the session is sent to the candidate network in the outsource request messages. Credentials that identify the UE are sent to the candidate network. The candidate network sends its credentials in the acknowledgement.

Service Outsourcing Execution Phase

Once the outsourcing SLA has been created, session establishment continues when the custodian network signals the UE to handover to the candidate access network. Mutual authentication between the UE and the candidate network is performed using credentials exchanged between the networks during the setup phase. The UE associates and performs a binding with the candidate network to allow session establishment. This

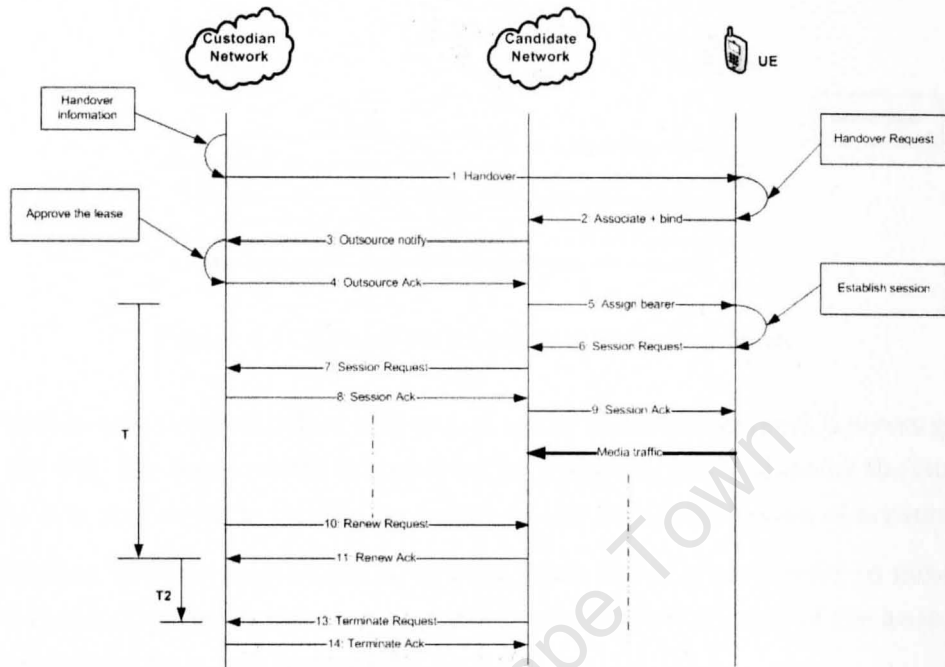


Figure 4.3: Transactions during the execution phase

process does not require input from the user. The candidate network creates and assigns a bearer with leased resources to the UE.

After the binding process signalling between the custodian network and the UE is routed via the candidate network. Both networks will monitor the use of resources for the lifetime of the session, which is limited by the lease period T . A new lease can be established when the current lease expires. During the lease period the custodian network performs continuous evaluation of resource availability on its access networks. It would be needless to renew the current lease if congestion levels subsided and resources became available. Figure 4.3 illustrates outsourcing transactions during the execution phase.

In the figure the candidate network notifies the custodian network (message 3) of the association and binding event performed by the UE. Moreover the session establishment requests (message 7) is relayed to the custodian network. In the case of local breakout media traffic can be confined to the candidate network. Alternatively media traffic can be tromboned to the custodian core network if the home routed option is enforced.

In the home routed option outgoing media traffic is routed to the custodian core network before it is forwarded to the destination. Likewise incoming media traffic will traverse

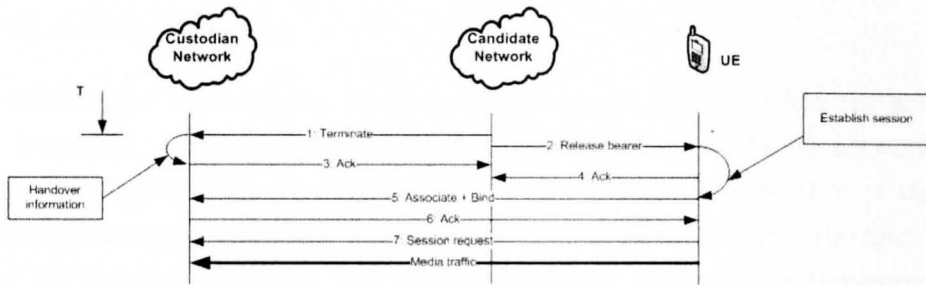


Figure 4.4: Service outsourcing termination phase

the custodian core network before it is routed to the candidate network's access gateway and to the UE. The home routed option is useful when charging is done by the custodian network. It is also useful in facilitating functions like legal interception of conversations. The custodian network may renew an existing lease before expiry (refer to message 10 in Fig 4.3). Monitoring of resource usage will continue upon renewal of the lease, whose lifetime is limited by a new period T_2 .

Service Outsourcing Termination Phase

A service outsourcing session can be terminated if the user terminates the media session before the expiry of a lease period. Moreover, an outsourcing session would be terminated at the expiry of an active lease period. In the termination phase, the candidate network sends a report on the usage of leased resources. The report contains the effective lease period and resource usage accounting information in applicable units, e.g., volume of traffic or duration. The bearer assigned to the UE is released to free the leased resources. This is illustrated in Fig. 4.4.

If the outsourcing service is terminated due to the expiry of the lease period T , the UE may be handed over to the custodian network. Once it associates and binds with the custodian network the active session is re-established; IP mobility mechanisms are required to accomplish the handover. However, application layer protocols, e.g., SIP, include session continuity mechanisms that may be used to accomplish the handover.

New and Active Services

Service outsourcing mitigates the blocking of new service requests when the access network is congested. Negotiations between the custodian and candidate networks occur before the new request can be admitted. The request would be blocked if the target access candidate network did not have the resources required to meet the QoS requirements of the session. Moreover, the request would be blocked if an outsourcing SLA could not be reached.

Service outsourcing for active sessions would be necessary if changes in custodian network conditions affected resource availability and compromised QoS commitments. Outsourcing could be performed for some user sessions to reduce the load on the custodian access network. This would prevent the need to drop sessions due to congestion.

Signalling Optimisation

Session establishment time has a great impact on user experience. Service outsourcing transactions involve the exchange of information between networks and the UE prior to session establishment. The response time of each network and the number of messages exchanged will directly influence the total session establishment delay. Service outsourcing transactions could be combined and sent in the same message to available networks; the information would be used once a pending decision is made. For example transaction 2 and 6 could be sent in the same message, while 5 and 9 could be combined (refer to Fig. 4.3). Moreover for network to network signalling transactions 3 and 7 could be merged, and 4 and 8 could be combined to reduce the session establishment delay. Signalling optimisation measures involving outsourcing of multiple user sessions are discussed in later sections.

4.2 Outsourcing Architecture

Service outsourcing utilises an architecture with a user and network agent. The user agent resides in the UE and provides network discovery information to the network agent. The network agent acquires and provisions network resources to meet user requests. Its role revolves around monitoring resource utilisation and conducting outsourcing interactions with candidate networks.

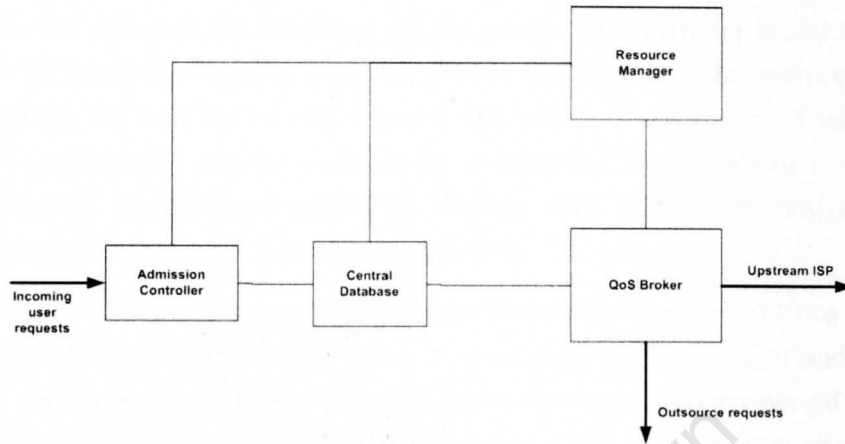


Figure 4.5: Outsourcing network agent

4.2.1 User Agent

Discovery of reachable access networks is performed by the user agent. Most networks periodically broadcast cell identity information, which the user agent receives and sends to the network agent. During session establishment the user agent sends a request to the custodian network containing updated information on reachable networks. During the service outsourcing execution phase the user agent manages the process of associating and binding to the candidate network. In the case of a UE with multiple interfaces that can be used simultaneously the user agent can maintain communication with the custodian network during and after the handover.

4.2.2 Network Agent

The network agent is the main entity in charge of the management and control of service outsourcing. It comprises of an admission controller, a resource manager and a QoS broker. Processing of user resource requests is done by the admission controller. The resource manager allocates and monitors the utilisation of network resources, i.e., local and external resources. External resources include leased resources and inter-domain upstream ISP connections. The QoS broker acquires and manages inter-domain resources.

The central database stores live network state information from the resource manager, QoS broker and admission controller. Figure 4.5 gives an illustration of the network

agent. Resource requests are processed by the admission controller based on resource availability information obtained from the central database and the resource manager. Session requests are admitted if the network has sufficient resources. If resource constraints are experienced, service outsourcing is triggered by the resource manager to lease the required capacity and admit the session. The QoS broker performs service outsourcing negotiations with candidate networks.

The database also stores information about service outsourcing transactions. The information is useful to the custodian network in visualising the acquisition and utilisation of network resources. The volume of leased resources during a given period would give an indication of the need for expanding the capacity of the network through a long term network upgrade. Moreover, the central database gives a view of the monetary cost of resource acquisition through service outsourcing.

4.2.3 Multiple Candidate Networks

In competitive scenarios there would be multiple networks from which a custodian network could lease resources. In this case an outsourcing request could be sent to multiple candidate networks. The outsource challenges and responses would contain different OP and T values to be used in the outsource SLA. These responses inherently form bids for the opportunity to provide the required service to the outsourced session. The custodian network would choose the candidate network that presents an opportunity for maximisation of benefits of outsourcing. This would be the network that agrees to an OP that allows the delivery of services at the required QoS level, and enables the custodian network to retain a substantial portion of user payments for the session.

4.2.4 Combined Outsourcing Transactions

Signalling optimisation in service outsourcing can be achieved by outsourcing multiple sessions in combined transactions. Using congestion thresholds to trigger service outsourcing, the custodian operator may perform outsourcing on multiple sessions that are considered to be less profitable. This may be done as a measure to ensure resource availability for more profitable higher profile sessions that would otherwise be blocked due to congestion. The combined outsourcing transactions are performed before the onset

of severe congestion. Fewer outsourcing messages would be exchanged between the custodian and candidate networks during SLA negotiation for a given number of sessions. This measure would reduce session establishment delay during service outsourcing.

4.3 Pricing of Network Resources

Network operators determine the cost of network resources and the charges to be levied for using the resources. Resource usage charges are meant to offset costs incurred in service delivery, and to generate profit for operators. In service outsourcing the cost of providing resources varies with different sessions. The custodian operator pays an agreed price to the candidate operator to lease network resources for a short period. The amount paid constitutes the outsourcing cost for the candidate operator. The user of the outsourced session pays a pre-determined price to the custodian operator. In order for the custodian operator to earn a profit from the outsourcing transaction, the pre-determined service charge (R_s) should be greater than the cost of outsourcing (C_o). The custodian operator's profit is given by

$$P_{HO} = R_{sH} - C_o > 0 \quad (4.1)$$

We make critical assumptions for modelling of service outsourcing scenarios. One major assumption made is that, the candidate operator would lease out requested resources at a price that permits the custodian operator to make a profit. Profit generation is also an important goal of the candidate operator. Thus, with respect to service outsourcing the price charged for leased resources (C_o) should offset the cost of provisioning such resources (C_r). Thus, the candidate operator's profit

$$P_{VO} = C_o - C_{rV} > 0 \quad (4.2)$$

C_o is the same for (4.1) and (4.2).

If the demand for resources on the custodian network were less than the resource capacity, network profits would be determined by the difference between network costs (CAPEX and OPEX) and service charges levied to users. Thus, $P_{HN} = R_{sH} - C_{rH}$; where P_{HN}

is the profit generated per user session without outsourcing, C_{rH} represents custodian network costs scaled down to the session, and R_{sH} is the service charge - refer to (4.1).

If the candidate operator did not lease resources to outsourced sessions, network profits would be determined by the extent to which revenue accruing from subscriber payments offsets costs of providing resource. Here $P_{VN} = R_{sV} - C_{rV}$; where P_{VN} is the candidate network profit for a user session, R_{sV} is the charge levied to users and C_{rV} is the cost of providing resources when abstracted to the session scale. C_{rV} is equivalent to the cost providing resources in (4.2).

The cost of providing resources varies from network to network. Moreover, it varies with time, influenced by OPEX and factors like interest rates on CAPEX loans. Operators acknowledge the importance of predictable service usage charges for their subscribers. Thus, service usage charges are set to be constant over long time frames. Industry regulation plays a major role in the determining the tariff rates for service usage. In practise the tariff rates used for standard services differ from network to network, and between different usage profiles on the same network. We consider a multimedia session that could be provided on the custodian or candidate network.

4.3.1 Custodian Network Outsource Profit

The custodian operator uses two service profiles with different levels of QoS guarantee in the pricing of multimedia services. Users choose a service profile that suits their communication needs. The profit margin generated by a session differs between service delivery on the custodian or candidate network. It is expected that higher profits would accrue from sessions delivered on the custodian network, i.e. $P_{HN} > P_{HO}$. It is also acceptable to assume that the cost of providing network resources is constant over short to mid-term time frames. Thus, the home service profit P_{HN} is constant over the same time frame. The service outsourcing profit P_{HO} would depend on the effective offer price. Both P_{HN} and P_{HO} are constant over the service outsourcing period T .

The lost profit opportunity (outsource cost margin - OCM) for an outsourced session is given by $P_{HN} - P_{HO}$. Thus,

$$OCM_H = C_o - C_{rH} \quad (4.3)$$

It is the aim of the custodian operator to minimise the value of OCM in order to retain a large portion of user payments. In addition to using low OP values in the outsourcing request, the custodian operator could use strategies like outsourcing of sessions belonging to users in lower service profiles, if lower C_0 values would apply.

4.3.2 Candidate Network Outsource Profit

The lost profit opportunity for the candidate network arises as a result of imposing levies that are different from normal levies imposed to home subscribers. Service outsourcing enables the candidate operator to utilise surplus network resources. Income from outsourced subscribers contributes in offsetting network costs and towards operator profit. The OCM for the candidate network revolves around the probability of blocking home subscriber sessions due to leasing resources to outsourced sessions, i.e. $P_{VN} - P_{VO}$.

$$OCM_V = R_{sV} - C_o \quad (4.4)$$

The candidate operator could minimise the OCM by challenging outsource offers with higher OP values.

4.3.3 Financial Implications of Service Outsourcing

The OCM gives an indication of the income operators involved in service outsourcing could lose. The custodian operator would be affected if, after outsourcing a session, resources became available on the access network. The resources would remain unused until the expiry of the current lease period. As mentioned above the candidate operator would be affected by the outsource cost margin if resource requests from home subscribers could not be met due to lack of resources, while some resources had been leased to outsourced sessions.

Using the following outsourcing parameters a numerical analysis of the income variation for the custodian and candidate networks is accessed. For the custodian network we set the maximum network capacity (number of sessions) at 20; the outsource OP is fixed at 1.5 monetary units. For the candidate network we set the maximum network capacity

at 30. The candidate network sets the negotiated outsource per-unit charge to a value greater than or equal to the offer price for a session (the value depends on the current network load). These parameters apply to both networks: all sessions require the same QoS level, the minimum duration of a session is 5 minutes and the duration for longer sessions is a multiple of 5 minutes. For simplicity we assume that all outsource requests are successful. Table C.1 in appendix C.1 provides the complete numerical data.

To achieve profitability the aim of both networks is to maintain the net income curve above the expected income cover. This is achieved by controlling income from service outsourcing with respect to income from serving more home sessions on either network. On the custodian network this would be achieved by not outsourcing sessions when there is available capacity, while for the candidate network it would be achieved by not blocking service requests from its subscribers due to hosting outsourced sessions. The candidate network would increase the negotiated per-unit price value whenever the load on its network increases; this would reduce the OCM in case requests from its subscribers were to be blocked. Figure 4.6 depicts the projected income variations; in Fig. 4.6 HN represents the custodian network, while VN represents the candidate network.

Several strategies aimed at securing a substantial share of shareable user payments can be adopted by the operators. The ultimate goal would be to minimise the extent to which an unfavourable OCM impacts network profit.

Operator attitude towards service outsourcing can be classified as optimistic or pessimistic.

An optimistic operator uses every available opportunity to participate in service outsourcing. On the other hand, a pessimistic operator does not participate in outsourcing considering the possibility that network conditions may change to make outsourcing unnecessary. A pessimistic custodian network holds session establishment requests anticipating that resources would become available when some active user sessions are ended. A pessimistic candidate operator rejects service outsourcing requests to reserve resources for future home subscriber requests.

Although the attitude of pessimistic operators may be justified as far as increasing chances of avoiding the OCM, it is clear that they reduce the chances of benefiting from service outsourcing. As mentioned earlier the benefits of service outsourcing include improved user satisfaction and prevention of churn for the custodian network. Candidate network benefits include utilisation of surplus resources, which improves network

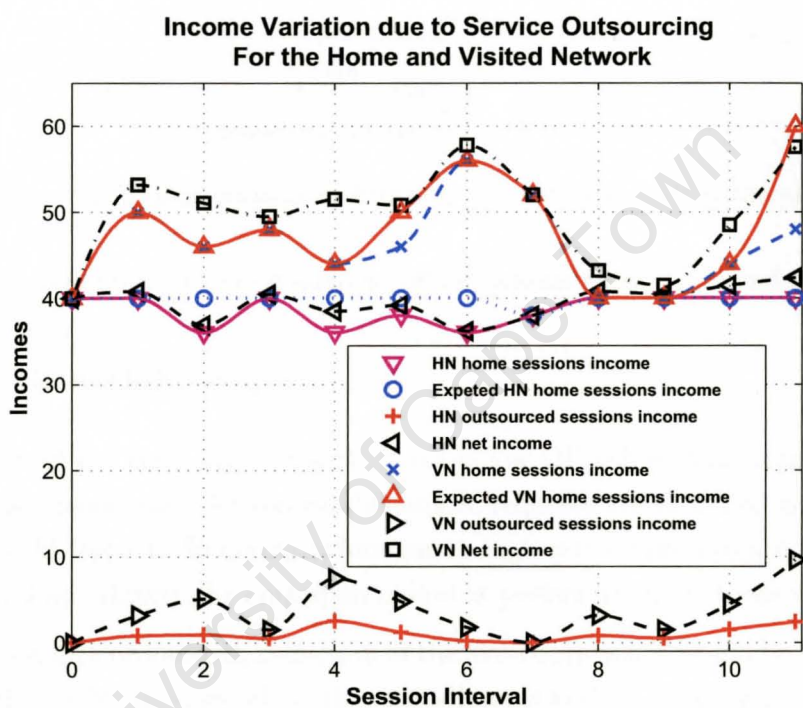


Figure 4.6: Income variation due to outsourcing

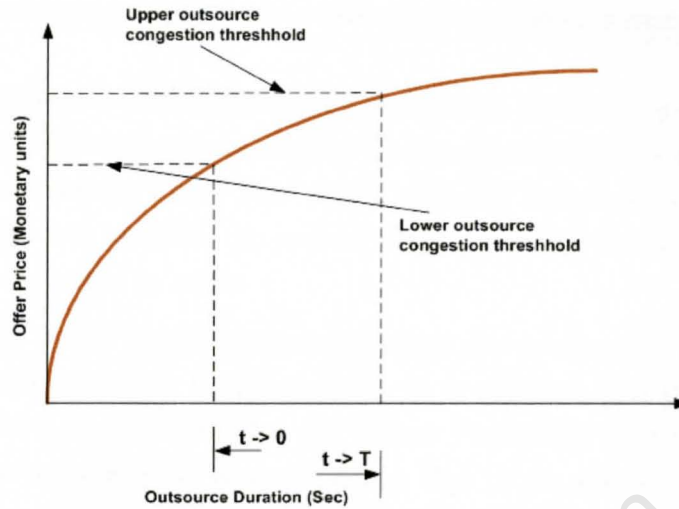


Figure 4.7: Sliding outsource duration (T) with congestion thresholds

efficiency and generates income needed to offset various network related costs.

Custodian Network Strategies

The main aim of the custodian network is to use low OP values, thus retaining a bigger portion of user payments. Session establishment requests are classified according to an estimated OCM impact. Requests belonging to high service profiles are likely to incur high OCM values. Hence, the custodian operator prefers to admit these requests.

If OP values cannot be lowered, reduction of the lease duration T would be more practical in limiting the OCM. It would allow the custodian network to regain early control of the session; if resources became free service delivery would resume on the custodian network. Generally the selection of an appropriate value of T should consider changes in network conditions. Congestion thresholds can be used to influence T . Longer values of T should be adopted when an upper congestion threshold is reached, while decreasing values of T are appropriate when congestion levels start dropping towards the lower outsource threshold. Figure 4.7 depicts the choice of T during service outsourcing.

In service outsourcing the custodian network does not reveal the actual price to be paid by users for services. Thus, the OP used for the first outsource request could be set low enough to greatly influence the OCM. This strategy relies on the candidate network basing its decisions on the OP received in outsource requests.

In addition to outsource duration strategies, pessimistic behaviour involving session holding can be initiated to delay outsourcing in anticipation of free resources becoming available. Service outsourcing would not occur if sufficient resources became available and the user attempted to establish the session. Resources could become available if active sessions were terminated by users.

Another strategy would involve outsourcing active low profile sessions to avail the QoS resources required for higher profile sessions. It is assumed that these sessions would be outsourced with comparatively low OP values. Practically more than one low profile session would be outsourced to accommodate a high profile session.

Candidate Network Strategies

The candidate network can minimise the OCM if OP values approximate to or exceed charges levied for resource usage by home subscribers. The candidate operator does not reveal the fee charged to home subscribers, thus the OP value set in the outsource challenge can be made to approximate these fees. However, if the challenge OP is set too high, there is a risk of termination of outsource SLA negotiations.

The length of the lease duration can be used to limit the duration over which an unfavourable OCM affects network profits. When the arrival of new resource requests leads to depletion of network resources, the strategy is to negotiate shorter outsource duration for leasing resources. This strategy facilitates the candidate network's participation in service outsourcing, especially when very low OP values are received in outsource requests. Thus, shorter outsource duration values are used instead of setting higher counter OP values.

An optimistic candidate operator may invest in the deployment of network technologies that avail abundant bandwidth. Using technology mechanisms that support service outsourcing, such a system will encourage competing operators to outsource bandwidth intensive services during peak demand periods. The candidate operator will benefit from income flows arising from service outsourcing.

4.4 Network Resource Supply and Demand Model

Bandwidth and buffer space are the main network transport resources available for selling to consumers. Generally the supply of these resources can be abstracted at the core and access network levels. At each level the total capacity is limited by physical media and switching capabilities. Resources in the network core are provisioned in aggregate volumes and sold over long time frames. This can be related to deploying wireless access networks to avail a certain amount of bandwidth through a process that involves long term investments in infrastructure and frequency spectrum. Assuming that frequency spectrum is sold on an open market through a bidding process, its demand will influence the cost.

Network operators are buyers and sellers of network resources; they buy in aggregate wholesale quantities and retail to end users. Using a network model composed of service users with active subscriptions to different network operators, the acquisition and retailing of resources by network operators can be abstracted. Each network operator owns a set of access networks that are connected to a core network to form an access domain (AD). The operators manage their network resources to provide adequate capacity to meet service contracts with their users. Each network operator connects and subscribes to network resources from upstream domains for providing inter-domain communications. The operators request a limited amount of inter-domain resources, which are provided and priced according to schemes specific to each domain. The inter-domain resources and the local resources in a given AD are priced by the network operator and used to serve users.

User Attitude to Pricing

Users buy network services from the access network operator and are interested in high quality services delivered at low costs. Their satisfaction is affected by service expectations defined in an SLA with the home operator (HO). The SLA may specify the properties of each service class to be used for each traffic session. In roaming environments users are considered to be less sensitive to the ownership of the physical network used to deliver the service, hence would accept any access network assigned to them. However, regardless of the network used, predictability of prices is an important requirement.

Time of the Day (ToD) pricing uses constant prices over a pre-defined period of time; hence it has the required predictability aspect. It is a model where the day is divided into two time frames, i.e., daytime and nighttime. Daytime prices are often set higher to counter network congestion due to expected high service demand; a reversed trend happens at night.

4.5 Conclusions

Service outsourcing would only be meaningful if participating networks could benefit financially. Moreover, a technical framework is required to support negotiations between operators to establish SLAs and perform handover for user sessions. Since outsourcing interactions involve competing operators, government regulations are required to prevent anti-competitive behaviour amongst operators. Without regulations, exploitation may occur if a candidate network levies unnecessarily high rates for outsourced services. The reason for the high charges might be to prevent the custodian network from utilising the benefits of service outsourcing, e.g., retention of subscribers that results from enhanced long-term user satisfaction.

In this chapter we presented mathematical illustrations of the conditions to be met for service outsourcing to be beneficial to the custodian and candidate operator. The custodian operator would benefit from service outsourcing by minimising the OP value. On the other hand the candidate operator benefits from high OP values. The various strategies for maximisation of service outsourcing benefits by either operator were explored.

Chapter 5

Service Outsourcing in IMS Frameworks

This chapter presents the design of a service outsourcing and charging architecture based on IMS.

5.1 IMS Service Outsourcing Architecture

The service outsourcing architecture consists of a home and a visited IMS domain consisting of IP connectivity access networks (IP-CAN) and an IP core. The architecture is illustrated in Fig. 5.1, where the user equipment (UE) attaches to the custodian (home) IP-CAN after entering the overlapping coverage area of the custodian and candidate networks.

The selection of the custodian IP-CAN is facilitated by information that is pre-configured in the UE to identify the P-CSCF to which SIP messages are routed. This information may also be assigned dynamically. The UE is assigned IP addresses and DNS configuration information during the process of attaching to the IP-CAN. This information can also be pre-configured in the UE. Once configured the UE can register with an IMS core by sending a register message.

Upon registration the UE can establish IMS sessions by sending an INVITE to the IMS core. Appendix D.1 presents a sample of an IMS registration transaction. If pre-conditions are specified in the INVITE the IMS peers will exchange three SDP offers

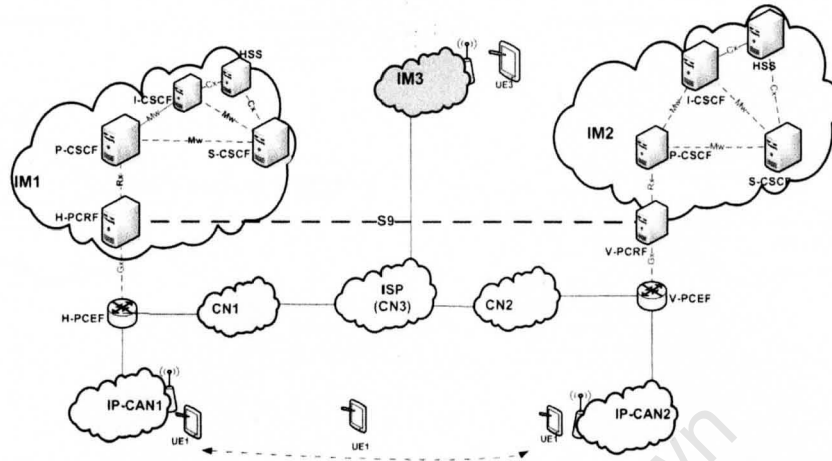


Figure 5.1: IMS service outsourcing architecture

and answers. For policy and charging control, the P-CSCF extracts session information and sends it to the Home-PCRF (H-PCRF) for creation of PCC rules [64]. Session establishment will be terminated if the required bandwidth is not available in the custodian IP-CAN. However, instead of terminating the establishment service outsourcing can be used to facilitate it by handing over the UE to the candidate IP-CAN where the required resources can be allocated.

Outsourcing in IMS requires a network agent resident in the core or instantiated as an application function in an application server as well as an outsourcing user agent resident in the IMS client. Interaction between the custodian and candidate networks can be achieved by using the S9 reference point standardised for PCC procedures [8]. Figure 5.2 illustrates the IMS outsourcing architecture used in this work. The custodian domain uses on the Home Network Agent (H-NA), while the candidate domain utilises the Visited Network Agent (V-NA).

5.1.1 Outsourcing Network Agent

The IMS IPTV service is used to illustrate the design of the service outsourcing framework. During the establishment of an IPTV session, when a SIP INVITE is received at the IPTV AS the application function retrieves service cost information from a tariff database in the billing domain (information on the design of the IPTV application server and the charging framework is presented in later sections). This information is used in

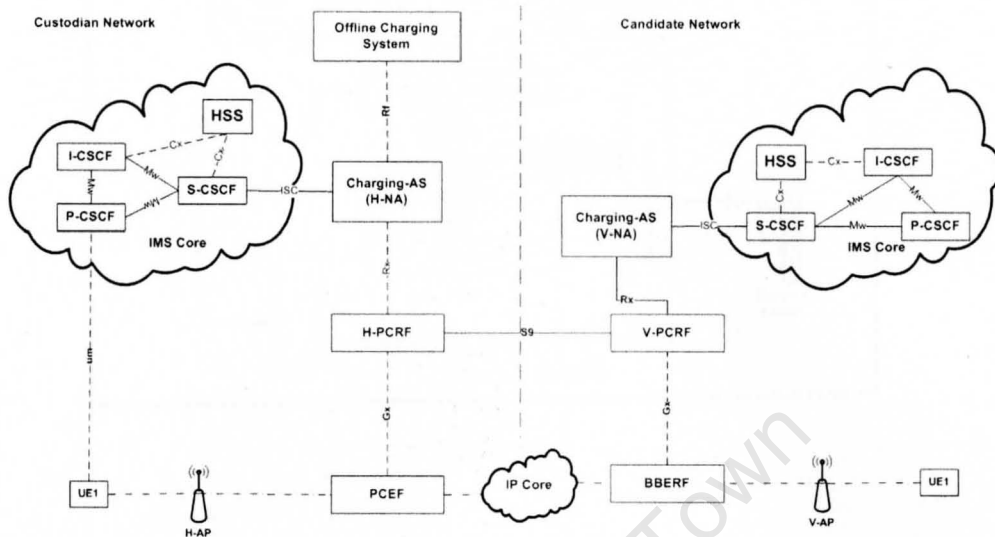


Figure 5.2: IMS outsourcing architecture

the outsourcing process performed by the H-NA, which communicates with the V-NA and the outsourcing user agent. The architecture of the H-NA is illustrated in Fig. 5.3.

During session establishment the process outsourcing function queries the resource utilisation database to determine the volume of committed network resources. This information is used to calculate the resource utilisation level, and subsequently the outsource offer price (OP) if a decision to outsource is made. If the resources requested for the IPTV session can be committed on the custodian IP-CAN the outsourcing process will be terminated and session control handed over to the IPTV AS. In the case where service outsourcing is to be performed the OP is used to create an outsourcing request that is sent to the candidate network.

The outsourcing request is sent to the candidate network by relaying across several interfaces including the Rx interface to the H-PCRF. It is then relayed to the V-PCRF using the S9 interface and finally to the V-NA using the Rx interface. If the request is accepted by the candidate network and a response received by the H-NA, it responds to the IPTV AS with an indication of the successful outsourcing operation. The IPTV AS sends an outsourcing command to the user agent, which performs a handover to the candidate network. In this design the outsource command is conveyed in a SIP 491 message that indicates a pending request. This will cause the client to de-register with the IMS core and initiate a handover to the candidate network.

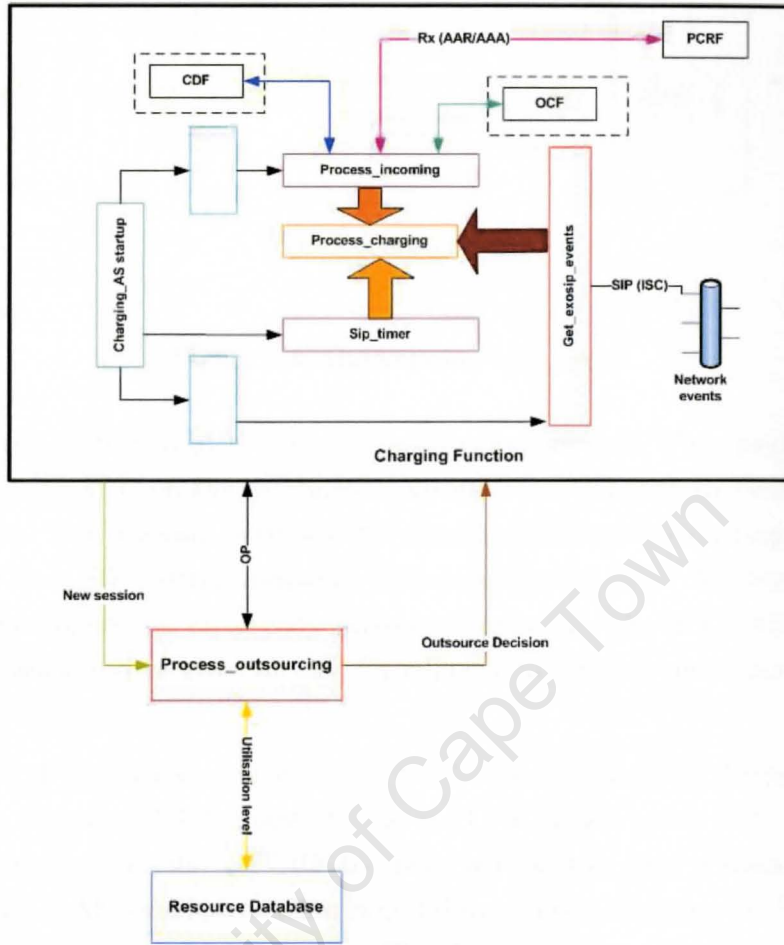


Figure 5.3: Architectural design of the network agent

5.1.2 Outsourcing User Agent

The outsourcing user agent receives and executes commands sent by the network agent. The user agent is integrated in the IMS client as shown in Fig. 5.4. To emulate an outsourcing handover, the IMS client (user agent) is configured as a multi-homed system with a primary connection to the custodian and a secondary connection to the candidate network. The primary connection is used to communicate with custodian network before outsourcing, while the secondary connection is activated upon receiving an outsourcing command.

During outsourcing, after successfully binding to the candidate network the IMS client performs a new registration with the IMS core using the candidate P-CSCF to route

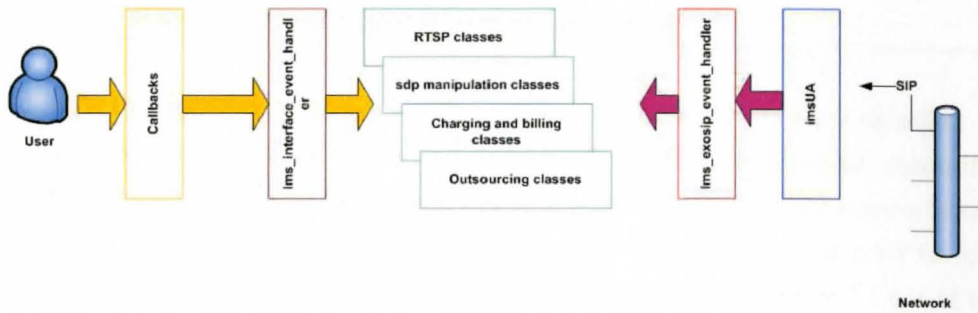


Figure 5.4: Outsourcing user agent

messages to the custodian IMS core and application servers. The client sends a SIP INVITE to the P-CSCF on the candidate network in an attempt to establish the outsourced session. For tracking purposes the client includes an outsourcing status in the SDP body of each SIP INVITE message. This prevents the IPTV AS from triggering a second outsourcing process on already outsourced requests. The IPTV AS resides in the custodian network and receives all SIP signalling relayed over custodian or candidate IP-CANs.

When the INVITE message is received the IPTV AS sends a Diameter Authentication and Authorisation Request (AA-Request) to the H-PCRF, which sends a Re-Authorisation Request (RA-request) to the V-PCRF to create a PCC rule that is installed at the V-PCEF. The IPTV AS completes session establishment upon receiving an Authentication and Authorisation Answer (AA-Answer). The duration of an active outsourced session would depend on the value of T in the outsource SLA; however, for simplicity purposes it remains active on the candidate network until terminated by the user.

5.1.3 Outsourcing Setup Phase

A binding between the SIP URI and the IP address of an IMS user agent is established during the registration process; the binding information is stored in the S-CSCF. It is noted that during service outsourcing the UE's IP address will change after a handover to a new access gateway. Thus, a new registration with the IMS domain is required to update the binding information upon handover to the candidate IP-CAN.

While attached to the custodian IP-CAN the user can establish IMS sessions with other IMS clients. For a VoIP session (refer to Fig. D.2 for VoIP session establishment SIP

signalling), the decision to outsource the session can be made by the custodian network during the exchange of SDP offers and answers if the H-PCRF cannot commit the required bandwidth in the IP-CAN. Bandwidth requirements and other session information contained in the SDP body of the INVITE (message 1 in Fig. 5.5) is used to formulate the outsource offer. The bandwidth information is conveyed in the `b=` lines of the SDP body. For outsourcing the OP and T values are derived by the network according criteria discussed in chapter 4.

The H-PCRF is an important entity for performing service outsourcing transactions in IMS communications. In normal operation the H-PCRF initiates PCC procedures after receiving session information extracted from SIP signalling by the P-CSCF (or an application function). During IMS session establishment with pre-conditions, the UPDATE (message 2) contains the final negotiated session parameters. Ideally the information sent to the H-PCRF is extracted from this message [64]. To accomplish this, the P-CSCF sends a Diameter AA-Request (message 3) to the H-PCRF to authorise the session. Figure 5.5 gives an illustration of service outsourcing procedures during IMS session establishment; only relevant SIP messages are shown. Note that if the requested bandwidth were available the H-PCRF would perform session binding and bearer authorisation, and send an AA-answer containing `DIAMETER_SUCCESS` to the P-CSCF.

During service outsourcing the H-PCRF creates a PCC rule specifying the requested bandwidth and charging information. This information is sent to the outsourcing H-NA (message 4), which creates an outsourcing request that is sent to the candidate network via the H-PCRF and S9 reference point (message 5). After establishing an outsourcing SLA between the networks, the H-PCRF sends an AA-Answer indicating the status of the outsourcing transaction to P-CSCF (message 6). The AA-Answer contains the candidate IP-CAN ID (identifier) and authorised media resources. This triggers the P-CSCF to send an outsource message to the calling UE to trigger a handover to the candidate IP-CAN (message 7).

5.1.4 Outsourcing Execution Phase

On receiving the outsource command, the UE de-registers (message 8) with the IMS core and performs a handover to the candidate network, after which it performs a new registration with the home IMS network using the new IP address (message 9). IMS

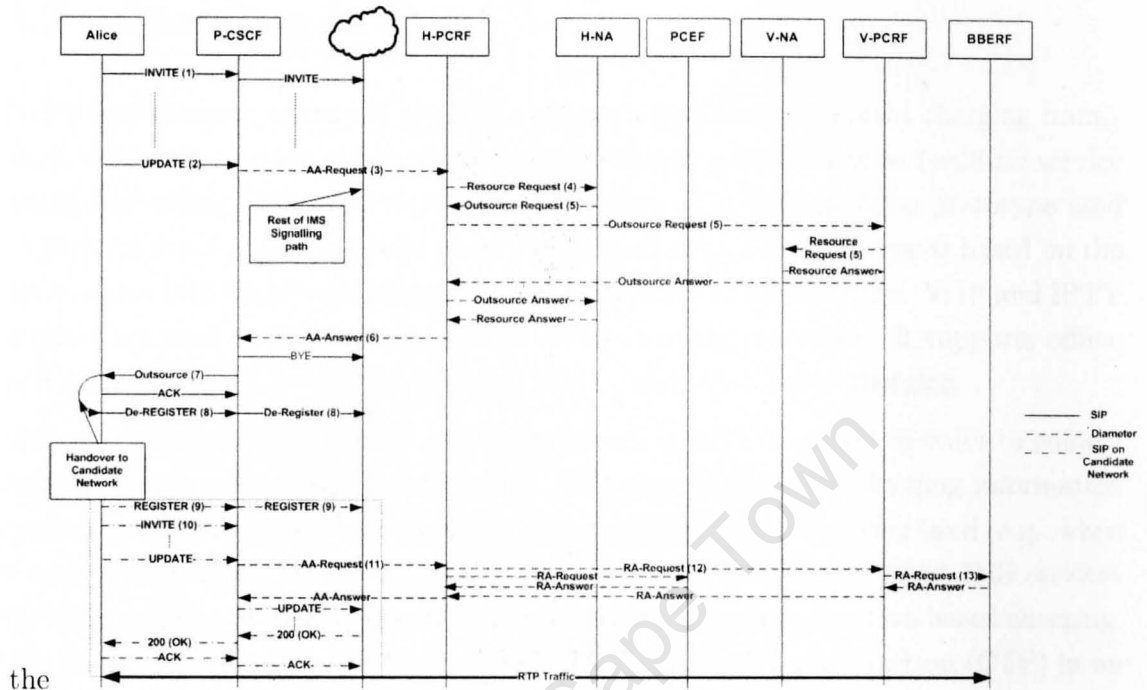


Figure 5.5: IMS session outsource negotiation

signalling traffic is now conveyed over the candidate IP-CAN. The UE then attempts to establish the session by sending an INVITE to the corresponding node (message 10). The INVITE contains the session authorisation information received in the outsource message. The SDP body is also populated with already negotiated session information, i.e. bandwidth and media codecs. Processing of the INVITE is performed according to standard IMS procedures. While performing PCC procedures, the P-CSCF sends an AA-Request to the H-PCRF (message 11), which sends a RA-Request to the V-PCRF to update the outsource request (message 12). This causes the V-PCRF to create and install PCC rules for the outsourced session at the BBERF or V-PCEF. Installation of PCC rules at the BBERF is done using a RA-Request (message 13).

Once session setup is completed RTP media traffic traverses the candidate IP-CAN. The candidate network can perform accounting and charging for the media session using the rules installed at the BBERF. In a home routed scenario, where media traffic traverses both candidate and custodian networks, the H-PCRF installs PCC rules at the H-PCEF to enable QoS and charging control. In this case the custodian network can perform flow level charging at the H-PCEF.

5.2 Charging in IMS

Policy and charging control in the IMS requires a standards compliant charging framework. Moreover, service outsourcing requires a charging framework to facilitate service rating and billing. This section presents the design of an IMS charging prototype used to perform proof of concept tests of service outsourcing. The prototype is based on the open source IMS framework released by the Fraunhofer Fokus Institute. VoIP and IPTV services are used to illustrate the design of the charging prototype. It supports offline and online charging, as well as duration, volume and event based charging.

QoS policy and charging control can be performed on all IMS planes in order to enforce resource usage control and charging [49]. The type of QoS and charging information generated at each plane depends on the communication scenario being used, e.g. when a user is connecting via the home or visited network. For session based IMS services the start and end of media delivery can be recorded to support duration based charging. The charging system requires the presence of a Charging Trigger Function (CTF) in an application server to detect the occurrence of chargeable events. The CTF interacts with the offline charging system using the Rf reference point and the online charging system across the Ro reference point.

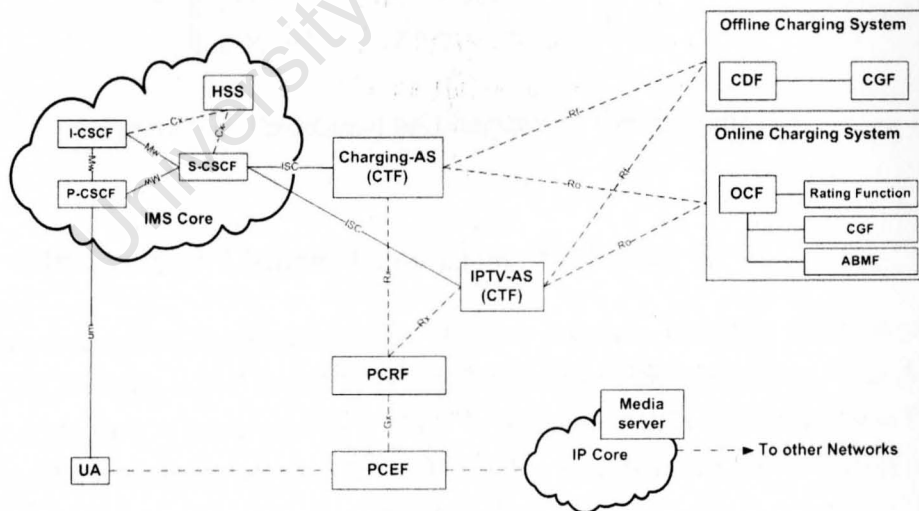


Figure 5.6: Charging system layout

Figure 5.6 illustrates the location of the IMS charging framework in the IMS architecture. The charging application server is integrated with a CTF to facilitate offline and online

charging. The design of the charging AS is shown in Fig. 5.7, in which SIP requests received by the `get_exosip_events` function are processed to determine if they contain chargeable events to trigger offline or online charging processes. The main chargeable events considered in this thesis are INVITE and BYE messages. These correspond to the start and end of IMS sessions.

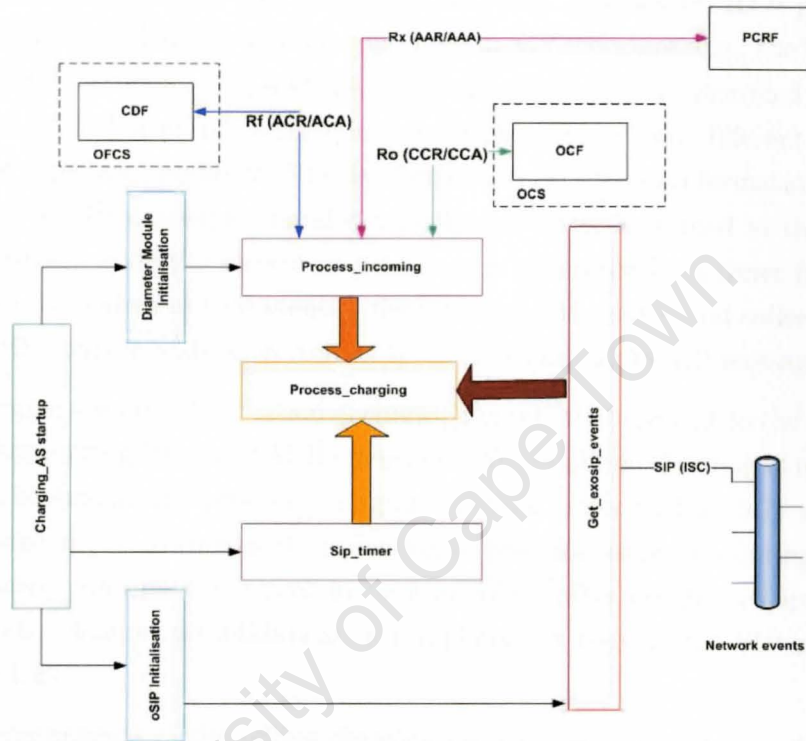


Figure 5.7: Functional architecture of the charging AS

5.2.1 Offline and Online Charging Systems

The offline and online charging systems utilise a common charging AS that connects to the S-CSCF using a SIP based IMS Service Control (ISC) interface. The AS has a Diameter node that is integrated as a CTF to interact with a Charging Data Function (CDF) for offline charging and an Online Charging Function (OCF) for online charging respectively.

The offline charging system requires specific information to process chargeable events. The information includes: a user ID, a unique session ID, an ID for the requested

service, media information (e.g., codecs and bandwidth), service request time and the IP addresses of the caller and callee UEs for peer-to-peer services. Similar information is required for online charging. The IMS public user identity (IMPU) can be used as the user ID to uniquely identify the user across domains. The session ID is generated from a combination of session information contained in SIP messages. e.g., the SIP session ID (cid) and the IMS contact (IP) address. The requested service is identified by the request URI for pre-defined application servers; peer-to-peer sessions use different request URIs. The service type, e.g., audio or video is identified from session information contained in SDP bodies of SIP messages. Bandwidth information is contained in the SDP bodies of SIP messages, while the service request time is generated by a timer function in the charging AS. IP addresses that identify the location of the caller and callee are extracted from the IMS contact address as well as IP-CAN name in the SIP message.

Offline charging involves the creation of charging events that are sent to the offline CDF in Diameter Accounting Request (ACR) messages. When offline charging is invoked during session establishment the `process_charging` function (refer to Fig. 5.7) authorises the requested session and initialises the collection of resource usage accounting information. The accounting information is sent in interim ACRs after pre-defined intervals e.g. 5 or 30 seconds. Session establishment is completed by routing the SIP request to the destination UE.

The main components of the offline charging system are a Diameter module and a Diameter message handler. The Diameter module listens for incoming ACR messages from application servers. Various AVPs are extracted from ACR messages and forwarded to the message handler for processing. The message handler receives charging information from the Diameter module. It performs billing and updates the user's account that is stored in a MySQL database. The structure of the offline charging system is illustrated in Fig. 5.8.

Offline charging starts by initialising the Diameter peer and listener processes. It receives incoming ACR messages, which are processed by the `Rf_ACA` function to retrieve AVPs and assign them to target variables. The AVPs include: charging ID, event charging data, duration charging data, volume charging data, nature of ACR (initial, interim or final) and subscriber identification information. Charging message handler functions are called from `Rf_ACA`, which also supplies the variables to be processed. A charging object is created for new charging sessions and events; subsequent ACR messages for session based

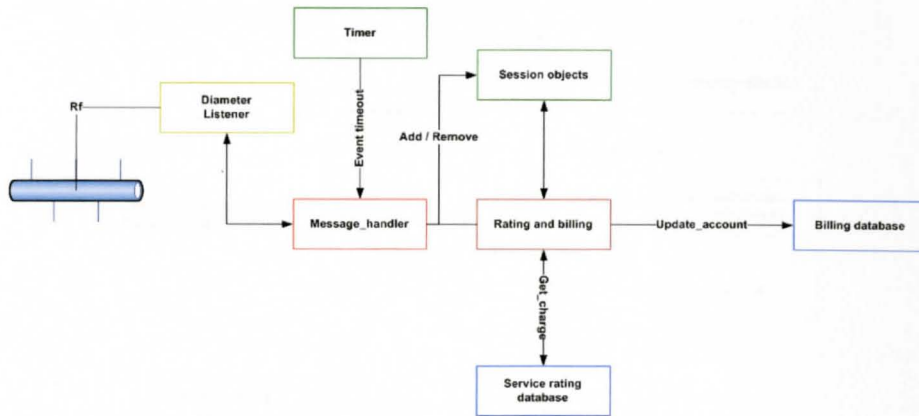


Figure 5.8: Structure of the offline charging system

charging will update the relevant object.

At session termination (when a SIP BYE method is received) a final ACR is sent to the CDF. Accounting Answer (ACA) messages from the offline charging system are received across the Rf reference point. For online charging Diameter credit control is supported across the Ro interface.

When online charging is invoked charging events are sent in Diameter Credit Control Request (CCR) messages to the OCF. The process_charging function creates an initial CCR and sends it to the OCS for authorisation. Upon receiving a Credit Control Answer (CCA) with allocated credits session authorisation is performed by routing the request to the destination UE. If no credits were allocated during credit control the request would be denied and the charging AS would send an error message, with the SIP code 402, indicating that payment is required.

The OCS comprises of a Diameter module, a charging event handler, a tariff module, an account balance management module and a subscriber billing database. The Diameter module is activated at start-up and listens for incoming CCR messages from the AS. The Ro_CCA function of the Diameter module extracts AVPs from CCR messages. The AVPs contain information that is used by the charging message handler to perform credit control. Initial and interim CCR messages cause the message handler to obtain service tariff information by invoking the tariff function. The amount of credits required to authorise a certain amount of resources is calculated using the tariff information. The account balance management function is used to check the availability of required credits. Figure 5.9 illustrates the functional architecture of the OCS.

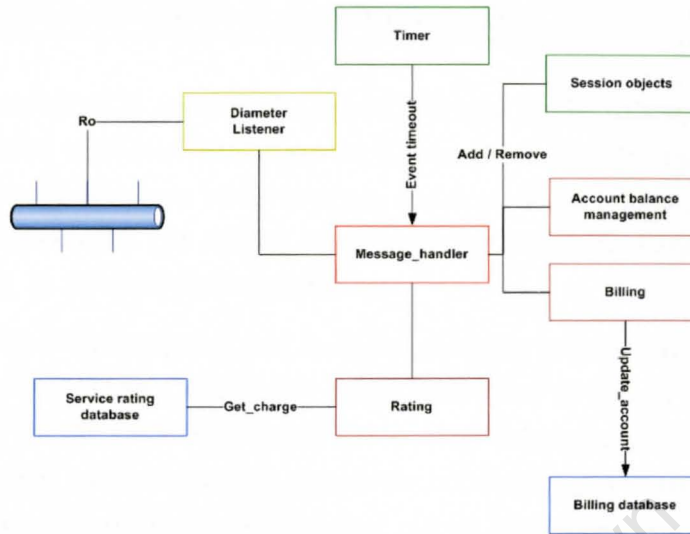


Figure 5.9: Functional architecture of the online charging system

Reserved credits are sent to the charging application function in AVPs contained in a CCA message. The implementation supports simultaneous reservation of credits for duration, volume and event charging. On receiving interim CCR messages the billing function updates the user's account by debiting used credits. When the final CCR is received at session termination unused reserved credits are credited to the user's account.

The process_charging function uses a timer to accomplish quota supervision on allocated credits. Quota supervision is achieved by real-time monitoring of the balance of allocated credits. Session termination is initiated when the credit balance is exhausted. At the end of pre-defined intervals the charging AS sends interim CCR to the OCS. In response the OCS allocates more credits for the session to ensure service continuity. A final CCR is send to the OCS when the session is terminated by the user or the charging AS. Online charging is achieved using the Ro_CCR function that creates Attribute Value Pairs (AVPs) and sends Diameter CCR messages to the OCS. Support for volume charging by the charging AS relies on an interface with the PCRF that is implemented by the Rx reference point.

Charging Events

The charging AF generates a charging event for each chargeable event that is then sent to respective charging systems for processing. As discussed earlier offline charging events

are sent in Diameter ACR messages to the CDF across the Rf reference point. The ACR message is created with standard Diameter headers that identify the AF and the CDF i.e., the origin and destination host, the origin and destination realms, and the accounting record type and number. For session based charging INTERIM charging events are sent periodically e.g. in 30 or 60 seconds intervals.

The Diameter Ro reference point used in online charging is based on the Diameter credit control application. When a chargeable event is detected at the AF a charging event is sent to the OCF in the OCS. This marks the start of a charging session for session based charging. For event based charging it signals the occurrence of a chargeable event. It was mentioned earlier that session based charging can be accomplished with unit reservation of credits, while event based charging can be achieved using immediate event charging and event charging with unit reservation (ECUR) [11].

In this framework unit based reservation of credits is used for session based charging; the number of reserved credits depends on service rating performed by the Rating Function (RF). Subsequent credits are allocated when INTERIM charging events are sent to the OCS. Event based charging is accomplished by ECUR. For session based charging, the Account Balance Management Function (ABMF) performs quota supervision to facilitate termination of service delivery when credit exhaustion occurs. An illustration of the quota supervision is given in Fig. 5.10. N represents the credit balance, n the number of reserved credits and y the number of used credits. Credits are allocated when the first CCR message is received and also when INTERIM CCR messages are received. Unused reserved credits are refunded to the subscriber's account at the end of the charging session. Typically the reception of a BYE message would mark the end of a charging session.

5.2.2 VoIP Charging System

The duration of a VoIP session cannot be pre-determined, thus session based charging is used for VoIP. To achieve this an initial Filter Criterion (iFC) is defined for VoIP services at the HSS to route SIP messages to the charging AS. The iFC is installed at the S-CSCF as part of the user profile when a user registers with the IMS domain. SIP messages including INVITE and BYE are routed to the charging AS (refer to Fig. 5.6). These messages constitute chargeable events, and they lead to the generation of charging

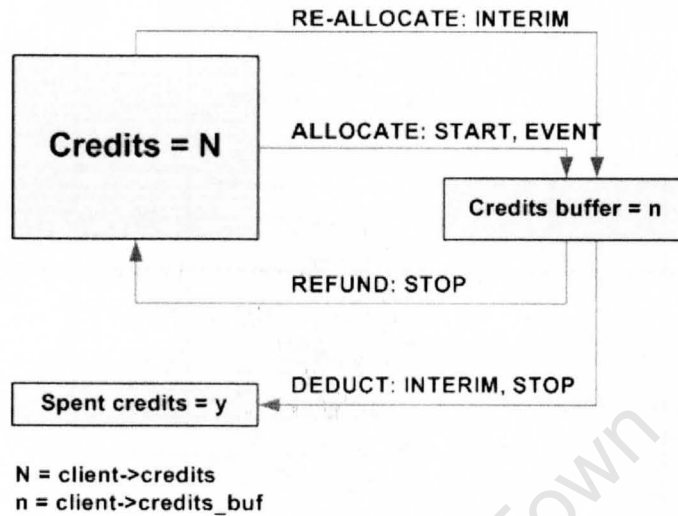


Figure 5.10: Online credit control

events that are sent to the OFCS or the OCS. An illustration of the signalling for VoIP online charging is given in Fig. 5.11. For clarity charging transactions are shown for the caller's IMS domain only; moreover, session description offer and answer exchanges are omitted.

The INVITE is intercepted and routed to the charging AS, where a CCR is generated and sent to the OCS. The OCS performs credit control and allocates a limited amount of credits to authorise the VoIP session. Upon receiving credits in the CCA the charging AS sends the INVITE message to the S-CSCF to continue session establishment. A 200OK is sent by UA2 when the INVITE is received and the call answered. This message traverses the same path as the INVITE. Quota supervision is performed by the OCS when INTERIM CCR messages are received. The VoIP session would be terminated if credits in the subscriber's account got exhausted. This would be accomplished by BYE messages being sent to participating users. If no credits were received in the CCA the charging AS would send an error message to UA1 indicating that payment was required.

The scenario depicted here involves duration charging for VoIP. The start of charging is marked by the reception of an ACK in response to the first 200OK. Offline charging signalling is similar in many aspects to online charging except the use of Diameter ACR and ACA messages that are sent to the CDF in the OCFS. Moreover, quota supervision is not performed for offline charging. Thus, session termination is not dependent on

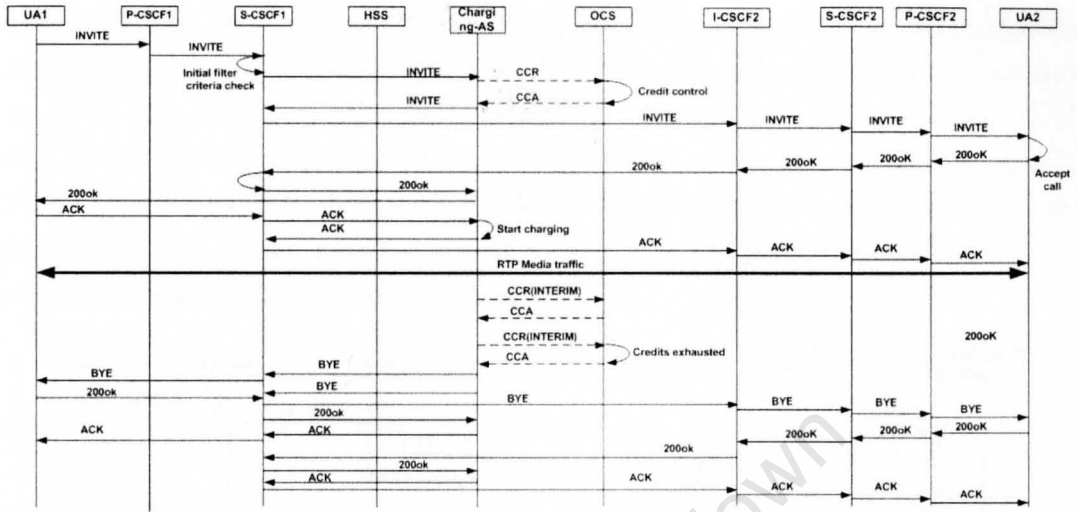


Figure 5.11: VoIP online charging

credit exhaustion. Note that in both offline and online charging session termination would occur if either user sent a BYE message.

5.2.3 IPTV Charging System

A CTF is integrated into the IPTV AS to detect the occurrence of chargeable events and send respective charging events to the OCFS or OCS. When a user requests an IPTV channel the S-CSCF routes SIP messages to the IPTV AS for the purpose of service delivery and charging. At the IPTV AS session processing is passed to the CTF for charging purposes. Offline and online charging procedures are handled similar to VoIP. Note that both session based and event based charging may be used for IPTV. Figure 5.12 illustrates signalling flow for IPTV charging.

In the IPTV system used in this thesis has the media control, service discovery and selection functions integrated into the IPTV AS. The IPTV AS is instantiated as a SIP UA and media delivery is performed by a third party RTP media server [13]. In the design only INVITE, BYE, 200OK and ACK messages are used to establish IPTV sessions.

The charging function in the IPTV AS supports duration and event charging. Volume charging is supported through interfacing with the PCRF and PCEF. When an INVITE arrives at the Charging AS it is processed to extract session information and identify

chargeable events. For every new session the charging process creates a charging session object identified by a charging ID. Moreover, a Diameter message is created and packaged with AVPs that contain the session information. The Diameter message is sent to the OFCS or OCS for offline or online charging respectively. For offline charging a Diameter ACR is sent to the OFCS, which sends back an ACA indicating if the request is authorised or not. For online charging a CCR is sent to the OCS, which reserves some credits and authorises the session for a given duration.

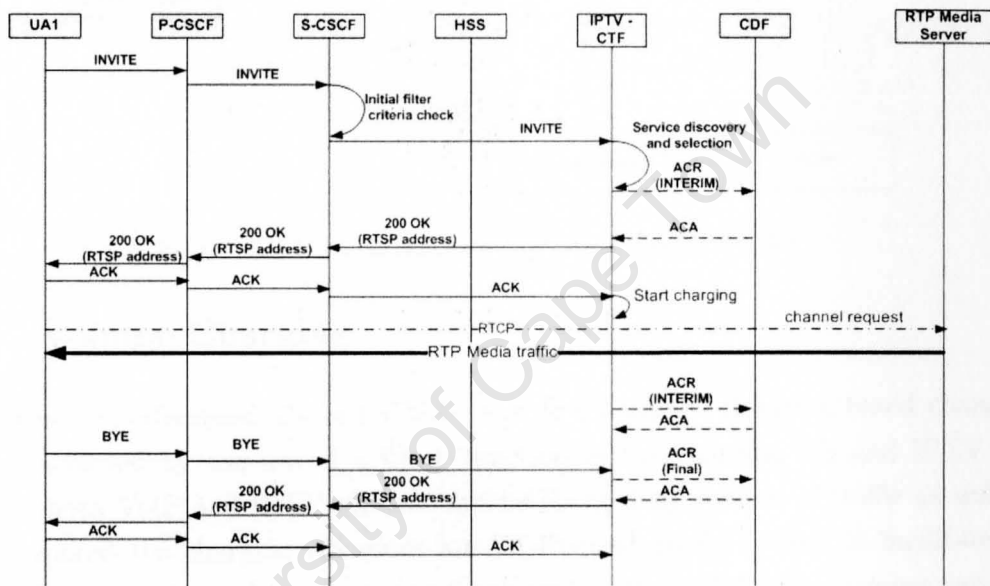


Figure 5.12: Charging for IPTV services

Charging and outsourcing are triggered when a TV channel request is received and processed by `get_exosip_events`. Charging is performed by a `charging_event_handler` function, which interacts with the OCS and OFCS to authorise the request. The charging function is implemented as an integrated CTF that builds Diameter messages to be sent to the OCS or the CDF. Offline charging transactions are handled by the `Rf_ACR` function, while online charging transactions are handled by `Ro_CCR`. The `outsourcing_event_handler` interacts with the PCRF to obtain resource utilisation information used to make outsourcing decisions. A detailed design architecture of the IPTV AS is illustrated in Fig. 5.13.

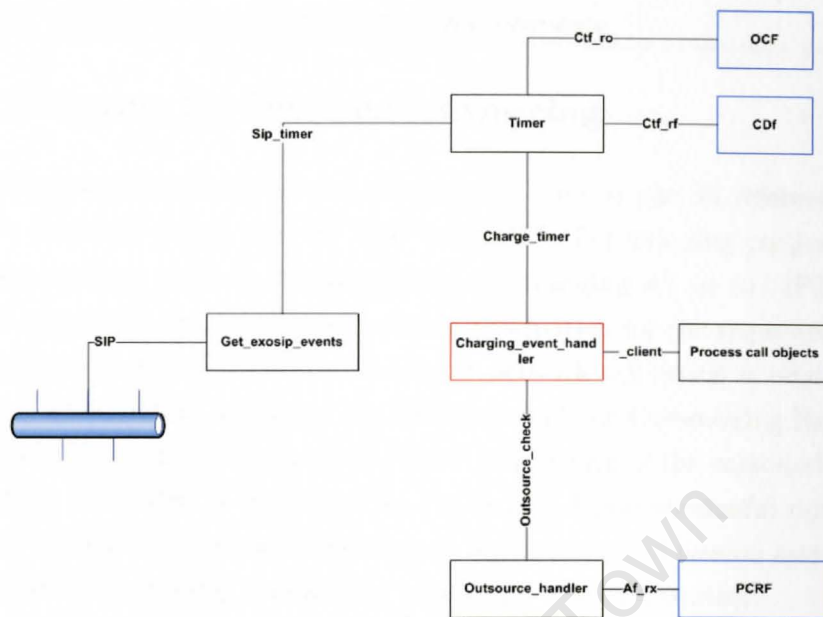


Figure 5.13: Architectural layout of the IPTV AS

5.2.4 Volume Charging

The mechanisms discussed above for VoIP and IPTV involve duration based charging, which is achieved by the use of a timer function in the charging AS and IPTV AS. However, both VoIP and IPTV services can be charged by volume of traffic as well. If this is required the charging AS sends an AA-Request to the PCRF to facilitate the creation of PCC rules to be installed and enforced at the PCEF. The request contains session information like the IP addresses of the calling and the called UEs. In this implementation the first AAR will trigger the start of volume charging, while the final AAR will terminate the charging process. Volume charging is triggered upon the success of the charging processes described above. AA-Answer messages received from the PCRF contain information on events occurring at the PCRF or those relayed from the PCEF.

The PCEF will perform accounting of all packets belonging to each flow identified by a given charging rule. Each rule is also identified by the charging ID received from the AS. Volume charging involves counting the bytes of RTP traffic that traverses the media plane. Accounting information is then sent to the OCS or OFCS via the PCRF and AS.

5.2.5 Charging for Service Outsourcing

Charging transactions during service outsourcing traverse the S9 reference point [8] in addition to reference points used for VoIP and IPTV. For charging purposes the service outsourcing network agent is integrated into the charging AS or the IPTV AS. When an INVITE is received the AS requests pricing information for the requested service from the billing domain. Information on the current network utilisation is retrieved from the H-NA. Once a decision to outsource the service is made an Outsourcing Request (OR) is sent to the candidate network via the H-PCRF. The design of the outsourcing framework with charging integrated is illustrated in Fig. 5.14¹. Upon successful outsourcing and handover of the UE to the candidate network, registration and session establishment are performed and subsequently followed by policy and charging control.

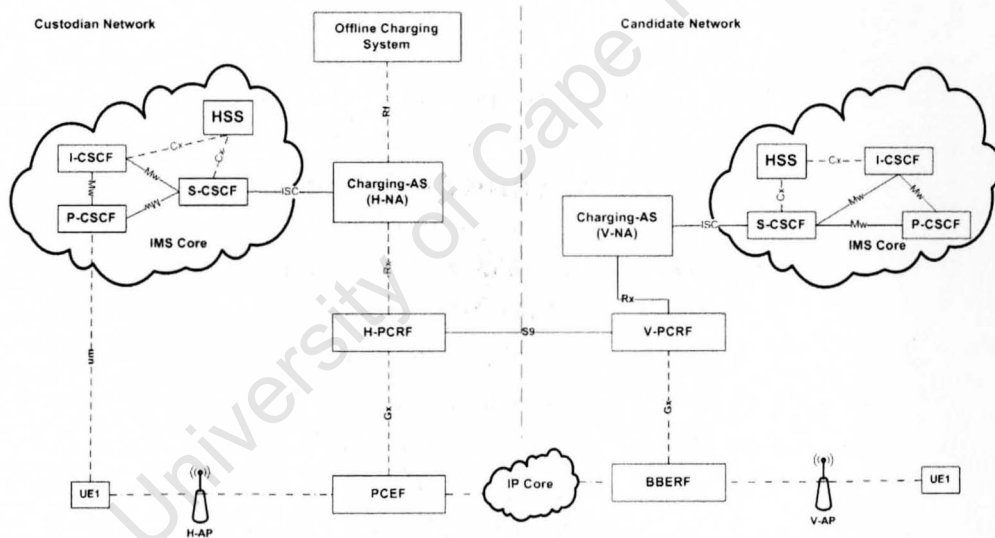


Figure 5.14: Outsource charging design layout

The H-NA is integrated into the charging AS to facilitate expedited outsourcing and charging transactions. Moreover, the Rx reference point is utilised for both charging control and outsourcing. A V-NA is present in the candidate network to perform outsourcing transactions for that network. Inter-domain information is relayed via the PCRFs to the outsourcing NAs in each domain.

¹This figure is repeated for the convenience of the reader

5.3 Summary

Service outsourcing in IMS utilises standard IMS transactions and PCC mechanisms. This chapter described the design of an IMS charging framework and a service outsourcing charging system. The IMS charging system presented uses INVITE and BYE messages as the main triggers of charging events. The design presented here is not exhaustive with respect to standards; however, it provides a useful framework for IMS research testbeds and is available for further development.

University of Cape Town

Chapter 6

Outsourcing Evaluation Framework

This chapter presents a service outsourcing evaluation framework based on IMS. The framework is used to perform tests in an environment that approximates to practical telecommunication scenarios. It includes a charging system developed in consideration of IMS standards released by the 3GPP [11]. It supports outsourcing requirements and provides an open source standards compliant prototype to the IMS research community. The source code is freely available to meet needs of any system that is designed according IMS standards, regardless of the programming language.

Open source projects are more suited for research due to easy customisation; moreover, using an evaluation framework in telecommunications research has specific advantages over simulation environments. They include generation of experimental results that closely measure to the expected performance of systems deployed in practical environments, where physical factors play a role. Framework entities can be distributed in different locations across the Internet to exploit changing network conditions; moreover, they can be subjected to stress testing. On the other hand simulation environments often require the selection of system properties like link delays, which influence performance results.

The charging and outsourcing framework is implemented to work with Fokus Open IMS testbeds, the UCT IPTV application server and the UCT IMS client. It is developed in the C programming language, and currently runs on Ubuntu Linux systems. The evaluation framework is used to perform a series of tests involving IPTV services. For this reason outsourcing is discussed with respect to IPTV services only.

6.1 Open Source IMS

The Open IMS Core (OSIMS) was developed and released by the Fraunhofer Fokus institute. It is a platform that enables the development and testing of functions and protocols related to IMS [9]. It implements IMS call session control functions (CSCF), i.e., P-CSCF, I-CSCF, S-CSCF and HSS. The CSCFs are programmed in C, while the HSS is implemented in Java and stores user information in a MySQL database. These entities interact using SIP based Mw and Diameter based Cx reference points. Figure 6.1 illustrates a simplified structure of the Fokus Open IMS core.

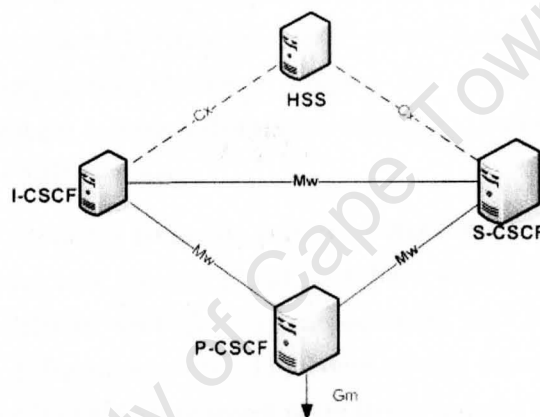


Figure 6.1: Layout of the Open IMS core

An OSIMS evaluation framework can be setup with different CSCFs collocated on the same server, alternatively they can be distributed on different systems. The OSIMS software used for the charging and outsourcing framework was installed on different computers running Ubuntu Linux [65] version 8.0.4 operating system. The decision to distribute IMS entities was to achieve an evaluation environment that approximates to practical network deployments. This allowed various tests performed to be influenced by network conditions in an Ethernet environment. Usage of the IMS testbed requires a domain naming system (DNS) server to resolve realm and entity names to respective IP addresses.

For outsourcing IMS home or custodian network realm names are specified when configuring a new installation of the Open IMS Core. The realm name is also set as the default visited domain for the network in accordance to IMS standards that specify the location of the P-CSCF to be resident in a visited [64].

6.2 UCT IMS Client

The UCT IMS client [66] is an implementation of a 3GPP SIP UA designed to work with the Open IMS core. It is a University of Cape Town initiative that was developed and released as open source. The client supports common IMS procedures, e.g. registration, VoIP session establishment, video calling and IPTV. The client is used extensively by the outsourcing framework. It is enhanced to support charging procedures used in outsourcing.

The UCT IMS client [13, 66] is implemented in C; it extensively uses the eXosip library to implement SIP methods. The architecture of the client is shown in Fig. 6.2¹. User actions, e.g. call establishment using the graphical interface are captured by callbacks functions. The actions are used by `ims_interface_event_handler`, which calls various functions, e.g. SDP manipulation classes to prepare SIP messages for sending to the IMS core. SDP manipulation is responsible for creating and populating the SDP body of SIP INVITE and other SIP messages.

The client also receives messages from the core for events such as incoming calls. The messages are processed by `ims_exosip_event_handler`, which passes the messages to functions in charge of different services, e.g., outsourcing and billing. The client can be extended to support different services. The client source information is presented in appendix E.1.1. This would enable interested users to participate in further development of outsourcing in IMS.

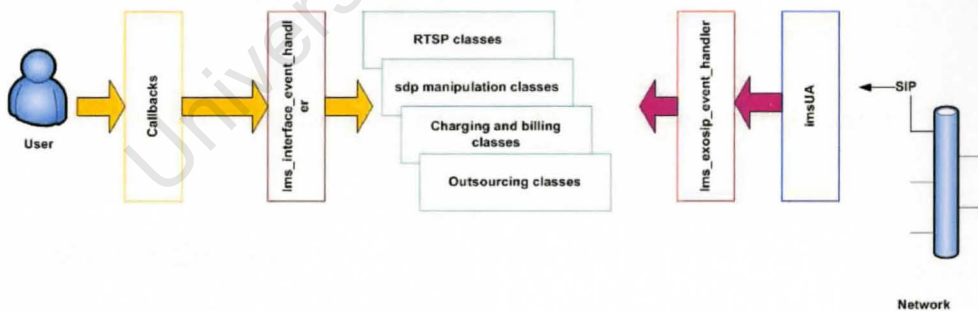


Figure 6.2: UCT IMS client

¹This figure is repeated for reader convenience

6.3 UCT IPTV Framework

The IPTV framework is a University of Cape Town initiative that was developed and released as open source [13]. It comprises an IPTV application server that is instantiated as a SIP user agent and integrated into the Fokus Open IMS framework. It is developed in the C programming language and utilises the eXosip library to handle SIP transactions; the architecture of IPTV AS is shown in Fig. 6.3. Once initialised the IPTV AS listens for incoming SIP requests from clients received via the core and processes them using `get_exosip_events`. The basic function of the IPTV AS is to fetch a media resource URL corresponding to a requested TV channel and send it in a SIP response to the client.

The IPTV framework is extended to support charging and outsourcing. The source code of the IPTV and charging framework is given in appendix E.1.2. IPTV channels are pre-coded in XML as key-value pairs and initialised during start-up. Sample entries for TV channels are presented in appendix E.1.2. More channels can be added by specifying a mapping between SIP request URIs and media resource URLs. Media URLs give the location of the resource on a media server.

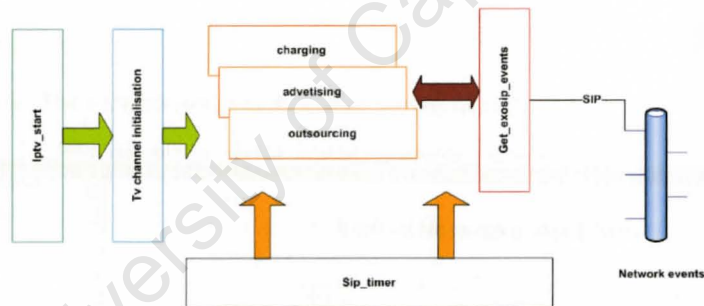


Figure 6.3: Functional structure of the IPTV application server

6.3.1 Media Server

Media resources are hosted on a Video LAN Codec (VLC) Real Time Streaming Protocol (RTSP) server that listens for incoming requests on port 5554. The requests are received directly from IMS clients as URLs of the form `rtsp://tv.imscore.ims:5554/channel1`, which corresponds to the value of the key in appendix E.1.2. The media server uses a configuration file that specifies the mapping between media channels and resources

like video files on the system. An example of the configuration file is given in appendix E.1.3. The location of the media server in the evaluation framework is shown in Fig. 6.8. Traffic from IMS clients connecting via custodian and candidate IP-CANs is routed via respective PCEFs.

6.3.2 HSS Configuration

The HSS is the IMS subscriber database that stores and manages network, service and user profiles; it is programmed in Java for OSIMS. In the evaluation framework the HSS provides a web based interface for configuring user profiles to add visited (candidate) networks used in outsourcing. Figure 6.4 illustrates addition of the `imsnet.ims` candidate network to the list of authorised visited networks. In Fig. 6.5 it is added to the profile of a user identified as Alice authorising IMS registration and session establishment via the candidate network.

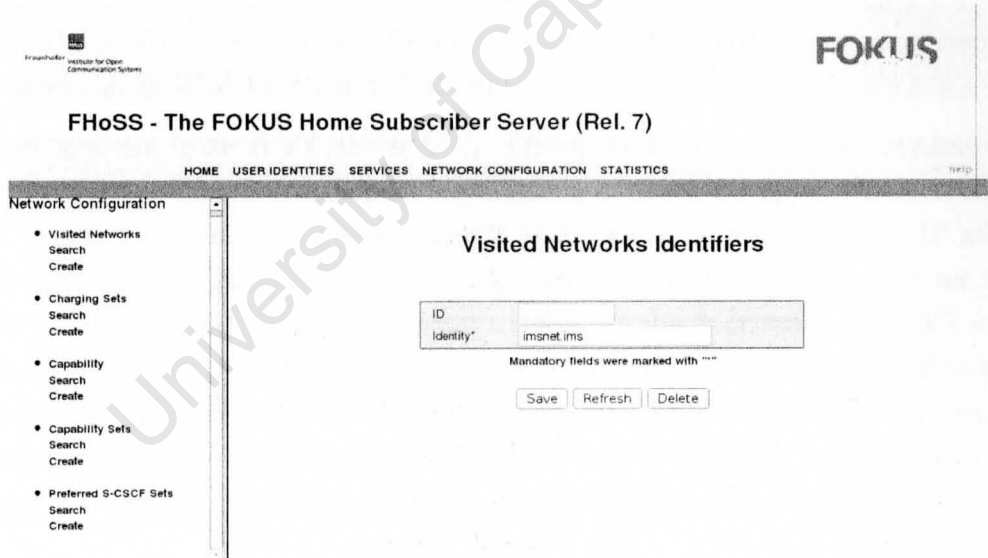


Figure 6.4: Provisioning the candidate network in IMS

HSS plays a key role in deploying and delivering services like IPTV using IMS control. The application server is attached to the IMS core using the HSS web configuration interface. This is accomplished by creating the application server, initial filter criteria (iFC), service point triggers (SPT) and shared iFC sets that are included in the service

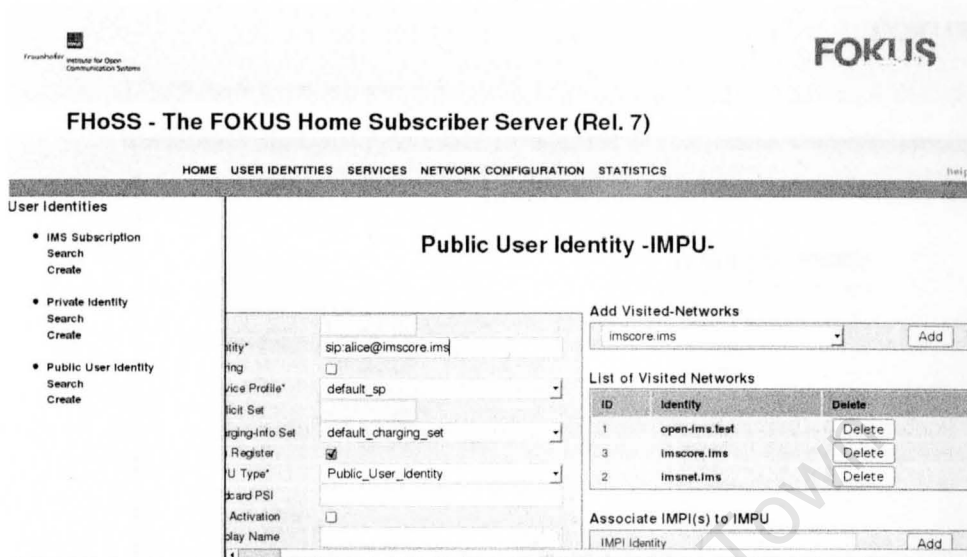


Figure 6.5: Authorisation of service outsourcing for IMS users

profile. The profile is used to identify and redirect matching SIP requests and responses arriving at the S-CSCF to the IPTV AS.

SIP messages sent to the IPTV AS use URIs of the form `channel_number@iptv.imscore.ims`, which identifies the requested channel and location of the IPTV application server. The server listens for incoming requests on port 8010, which together with its IP address are configured at the HSS; Fig. 6.6 illustrates definition of the application server, while Fig. 6.7 illustrates creation of service point triggers. In this framework the IPTV service profile is attached to the default service profile and downloaded to the S-CSCF to be associated with each user upon registering with the core. For charging and outsourcing the iFC causes all SIP messages related to IPTV to be sent to the IPTV AS. Appendix E.1.2 presents more HSS configurations showing the iFC and service profiles.

6.4 Outsourcing Framework

The evaluation framework emulates two IMS domains for custodian and candidate networks as illustrated in Fig. 6.8. All entities in the framework are installed on Intel dual core computers running the Ubuntu Linux operating system. IP connectivity for IMS core entities like CSCFs and HSS is provided by an Ethernet based network with a DNS

FOKUS Home Subscriber Server (Rel. 7)

IE USER IDENTITIES SERVICES NETWORK CONFIGURATION STATISTICS

Application Server -AS-

SH Interface - Permissions

Permission for	UDR	PUR	SNR
Allowed	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Request	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Repository-Data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
IMPU	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RMS User State	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S-CSCF Name	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
IFC	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Location	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
User-State	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Charging-Info	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
MS-ASN	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PSI Activation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
DSAI	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Always-Rep-Data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Mandatory fields were marked with *

Save Refresh

ID	
Name*	IPV
Server Name*	sp.10.20.0.1.80.10
Domain FQDN*	iptv.imscore.ims
Default Handling*	Session - Continued
Service Info	
Rep-Data Limit	1024

Figure 6.6: Definition of IPTV application server at the HSS

server. The choice of an Ubuntu Linux platform was based on the platform supported by Fokus OSIMS. Moreover, Linux is a free open source platform with support for traffic control and IP accounting. These functions are required for volume charging modules.

The specifications of the computers include 2Gbytes of RAM, 2.7GHz processors and 1Gbps Ethernet network interfaces. The control and media planes are physically separated at the system hosting each PCEF. This system has 3 Ethernet cards to create sub-networks for IP access, IMS core and IP core networks by utilising the Linux TCP/IP stack and routing module.

The UCT IMS client is deployed on a computer with two Ethernet interfaces. The primary interface is used to connect to the custodian IP-CAN, while the secondary interface connects to the candidate IP-CAN and is activated during the outsourcing execution phase. In total 4 computers are used to run multiple instances of the client in the testing stage.

FOKUS

Home Subscriber Server (Rel. 7)

S SERVICES NETWORK CONFIGURATION STATISTICS

Trigger Point -TP-

<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%;">ID</td> <td style="width: 80%;"></td> </tr> <tr> <td>Name*</td> <td>IPTV_trigger</td> </tr> <tr> <td>Condition Type CNF*</td> <td>Conjunctive Normal Format</td> </tr> </table> <p style="font-size: small; text-align: center;">Mandatory fields were marked with ***</p> <p style="text-align: center;"> <input type="button" value="Save"/> <input type="button" value="Refresh"/> <input type="button" value="Delete"/> </p>	ID		Name*	IPTV_trigger	Condition Type CNF*	Conjunctive Normal Format	<table style="width: 100%; border-collapse: collapse;"> <tr> <td colspan="2">Attach IFC</td> </tr> <tr> <td>Select IFC</td> <td><input type="button" value="Attach"/></td> </tr> </table> <table style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <tr> <td colspan="3" style="text-align: center; font-size: small;">List of attached IFCs</td> </tr> <tr> <td style="width: 10%; text-align: center;">ID</td> <td style="width: 70%; text-align: center;">IFC Name</td> <td style="width: 20%; text-align: center;">Detach</td> </tr> <tr> <td style="text-align: center;">2</td> <td style="text-align: center;">iptv_filter</td> <td style="text-align: center;"><input type="button" value="Detach"/></td> </tr> </table>	Attach IFC		Select IFC	<input type="button" value="Attach"/>	List of attached IFCs			ID	IFC Name	Detach	2	iptv_filter	<input type="button" value="Detach"/>
ID																				
Name*	IPTV_trigger																			
Condition Type CNF*	Conjunctive Normal Format																			
Attach IFC																				
Select IFC	<input type="button" value="Attach"/>																			
List of attached IFCs																				
ID	IFC Name	Detach																		
2	iptv_filter	<input type="button" value="Detach"/>																		

Add SPTs to Trigger Point

Not	<input type="checkbox"/>	SIP Method	INVITE	<input type="button" value="Delete"/>
OR				
Request-URI				
AND				
Not	<input type="checkbox"/>	SIP Header	To	<input type="button" value="Delete"/>
		SIP Header Content	'iptv.imscore.ims'	
OR				
Request-URI				
AND				
Request-URI				

Figure 6.7: Definition of service point triggers for IPTV

6.4.1 Outsourcing User Agent

Basic custodian network configuration information including the name of the P-CSCF, the IMS realm name together with user credential are encoded in the IMS client to enable registration and session establishment. For outsourcing evaluation purposes the IPTV request URI sent to the application server is also configured in the client. On receiving responses from the IPTV server the client identifies outsourcing commands by inspecting the content header in SIP messages for the pattern `OCN="imsnet.ims"`. This information is conveyed in 'request pending' or SIP 491 messages processed by `ims_exosip_event_handler` (refer to Fig. 6.2) to call outsourcing classes.

If outsourcing is required the agent triggers the IMS client to perform a de-registration from the custodian network, and then performs subsequent registration with the candidate network and finally a new attempt to re-establish the original session is performed. This is performed by registering with the candidate network during the execution phase. The outsourcing agent updates the address of the P-CSCF to match candidate network settings, including changing the IMS client's SIP contact address to the IP address of

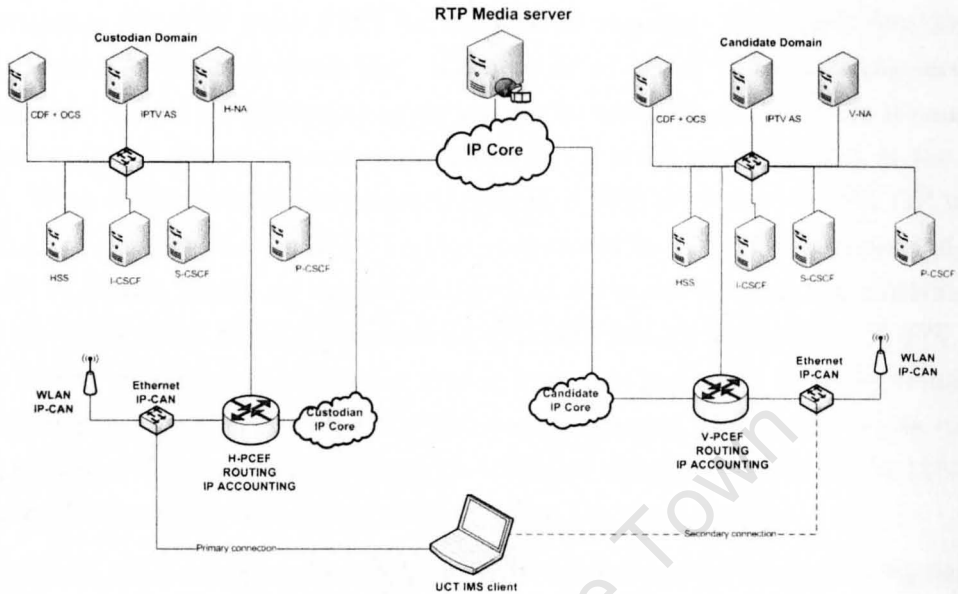


Figure 6.8: Layout of the evaluation framework

the secondary Ethernet interface. As presented in earlier chapters signalling and media traffic will traverse the candidate IP-CAN after outsourcing.

The outsourcing agent is integrated into the UCT IMS client. The source code for the outsourcing agent is presented as part of the UCT IMS client in appendix E.1.4 and released to the open source community.

6.4.2 Outsourcing network agent

The network agent is implemented as an application function and integrated in the IPTV application server. In the evaluation platform it tracks the number of active IPTV sessions admitted via custodian and candidate networks. This information is stored in a resource utilisation database and used to compute utilisation levels of each network and influence outsourcing decisions. The maximum capacity of the custodian network is fixed at 7 IPTV sessions. Outsourcing is performed on all session establishment requests that arrive when this capacity is consumed. The choice of 7 sessions for the custodian network ensured that it served at least twice the number of admissible active sessions (the virtual capacity) used during the testing stage.

The Network Agent (NA) in the custodian network has an outsourcing tariff function

that computes the offer price (OP) for outsourced sessions. The tariff function uses virtual capacity utilisation levels (i.e. the number of active IPTV sessions served on custodian and candidate networks) as an input into computing the OP. It is computed as a percentage of virtual capacity utilisation and a pre-determined cost of the IPTV service. Thus, if the resource utilisation threshold is 80% (or 8 sessions) the OP value is set to $0.8 * P$, where P is the IPTV service cost stored in a MySQL database. An 80% threshold is chosen based on the knowledge that networks become less efficient when utilisation levels reach 80% of the capacity. The OP has an upper limit of 90% of the service cost to guarantee some revenue arising from user payments would be retained by the custodian network. The outsource duration is changed to different values for each testing scenario. A timed countdown on the effective duration is enforced by the NA at session establishment to terminated sessions when the timer expires.

The candidate NA is configured to accept any OP values received in outsourcing requests. The custodian NA sends a maximum of 3 active outsourcing requests to achieve a virtual capacity of 10 IPTV sessions that can be admitted on both networks. The custodian NA rejects all session establishment requests that are received after the virtual capacity is exhausted. Appendix E.1.5 presents the source code for the NA, which is available as part of the IPTV application server.

The NA and IPTV AS use three types of SIP message to communicate outsourcing decisions to the IMS client. When an IPTV session is admitted to the the custodian network a 200Ok message is sent to the client. This message also carries the URL (for example URL="rtsp://tv.imscore.ims/channel1") of the requested media channel in its content header. As mentioned earlier a message containing the SIP code 491 is sent when outsourcing is performed: in addition to the media URL this message also carries the name of the candidate network (for example the content would contain OCN="open-ims.test" URL="rtsp://tv.imscore.ims/channel1"). On the other hand a SIP 480 message is sent to the client in the event of failing to admit or outsource a session due to exhaustion of the virtual capacity. It indicates unavailability of the requested resource.

6.4.3 Charging Framework

The charging framework comprises a charging trigger function, an online charging system, an offline charging system, a PCRF and PCEF. The IMS charging trigger func-

tion (CTF) is integrated into the IPTV AS to support online and offline charging. The framework uses SIP INVITE and BYE messages to initiate and terminate charging. Thus, INVITE and BYE messages constitute primary chargeable events. Like the IPTV AS the CTF is programmed in C. It implements duration charging by starting a timer that is incremented for each second during the lifetime of an active IPTV session. The CTF supports multiple concurrent IPTV sessions from different users by initialising a charging session for each request. Charging sessions are uniquely distinguished using session identifiers generated from the value of SIP connection IDs or 'cid' that are assigned allocated by the IPTV AS for IPTV sessions.

The CTF performs charging after enforcing service (credit) authorisation for online charging. Service authorisation is achieved by returning a success status to IPTV AS SIP functions. If the service is not authorised for example as a result of insufficient credits the IPTV AS sends a SIP message with code 402 to indicate payment requirements to the user.

In the cause of an active session, and after pre-defined charging interim periods for example 5 to 30 seconds the CTF sends information on duration of usage to the online charging function (OCF) for online charging or the charging data function (CDF) for offline charging. The online charging system replies with new credit allocations to authorise further service delivery. The CTF triggers the IPTV AS to terminate an active session by sending a BYE if the amount of authorised credits are exhausted before the expiry of an interim duration. For this reason interim periods are configured to be shorter than the amount of initial credits allocated at session activation.

The Ro interface between the IPTV AS and the online charging system is implemented in C and utilises the C Diameter Peer (CDP) package from Fokus Institute. New Diameter attribute value pairs (AVP) and applications are defined for credit control across the Ro interface². The AVPs are used to convey charging detail records (CDR) for duration, event and volume charging. A customised Diameter message structure is defined for these charging categories; each message transports three CDR AVPs for duration, volume and event based charging, with only the applicable charging category containing usable values. The Rf interface used for offline charging is implemented using the CDP package as well. However, it does not perform credit control but re-uses the AVPs defined for online charging; an offline Diameter application is defined. The source code for the

²AVP and application codes used in this framework are custom and not standardised

CTF is available in appendix E.1.6 and has been released as open source. It presents an implementation of the Rf and Ro interfaces. Figure 6.9 illustrates the layout of the charging framework.

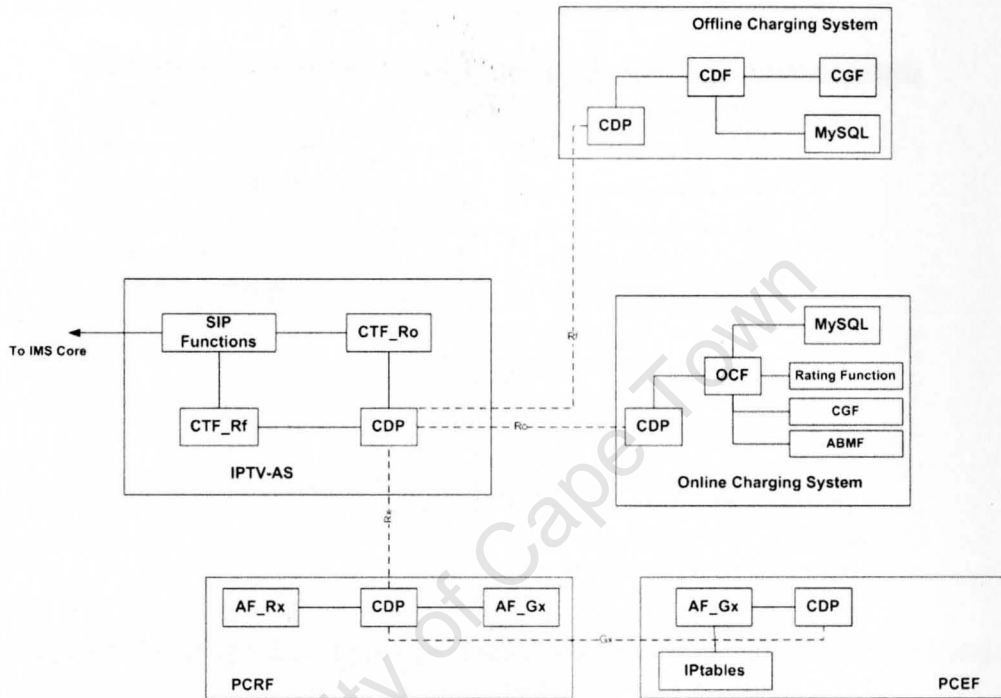


Figure 6.9: Charging Framework

Volume based charging is implemented through an Rx interface with the PCRF. For IPTV video on demand services the CTF sends the IP address of the requesting IMS client, the IP address of the media server and the charging identifier. The information is used by the PCRF to create IP filter rules that are eventually sent to the PCEF across the Gx interface. The Fokus CDP package is used for implementing Rx and Gx interfaces. Custom Diameter applications and AVPs are created for each interface to convey CDRs to the online or offline charging system via the CTF.

Communication between the CTF, OCF, CDF and PCRF is achieved by Diameter peers co-located with each function. Bindings are established between the peers using respective IP addresses and port numbers defined in configuration files. A sample Diameter peer configuration file that is used by the CTF to establish bindings with the OCF and CDF is given in appendix E.1.2.

The charging framework can invoke online or offline charging depending on user profile information stored in a MySQL database. Alternatively user input made through the IMS client and conveyed in SIP INVITE messages can be used. Figure 6.10 illustrates selection of online or offline charging at the IMS client³.

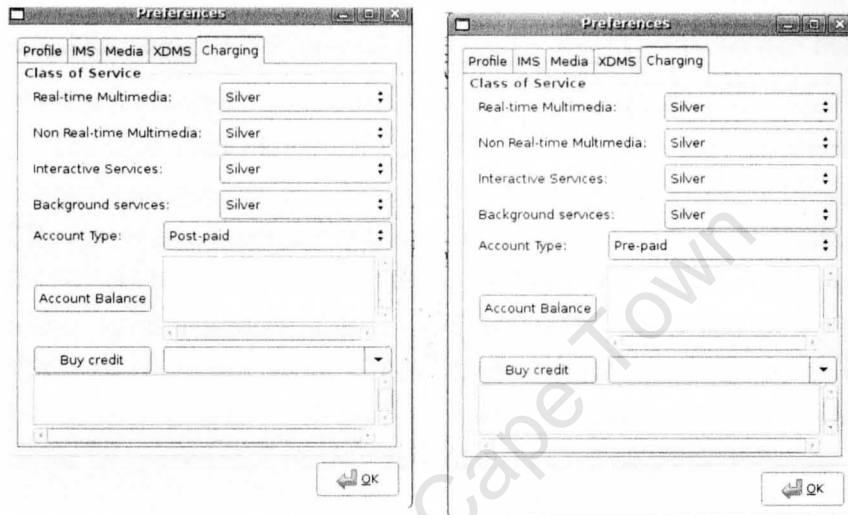


Figure 6.10: Prepaid and post paid selections for online and offline charging

Online Charging System

The online charging system (OCS) is programmed in C and utilises the Fokus CDP to implement credit control with the IPTV AS, where the CTF is located. Charging events generated by the CTF are received in Credit Control Request (CCR) messages, conveyed in AVPs. CCR messages marking the start of a charging session, e.g., duration based contain charging action AVPs with the value 'START'. This is triggered by INVITE messages; BYE messages convey the value 'STOP', interim CCRs contain 'INTERIM', while a CCR for an event based service conveys 'EVENT'.

Service rating is performed by searching the services table in the MySQL database for the price corresponding to the requested TV channel. Figure 6.11 illustrates service rating information for IPTV channel 1. This process determines the amount to be debited on a subscriber's account identified by the IMS public ID (IMPU) received in an AVP

³Note that it is denoted as pre-paid and post paid charging for simplicity and user convenience

created by the CTF. Event based charging is performed by reservation of units equivalent to the corresponding charge. Session based charging involves reservation of credit units required to deliver the IPTV service for 60 seconds, i.e., 60 units. Furthermore, volume charging credits can be allocated. Half the amount of credits are allocated if the balance on the user's account is less than 60 credits; the halving factor is used until the balance is exhausted. Allocated credits are sent to the CTF to authorise service delivery.

Account balances are stored in a balances table in the MySQL database; this is illustrated in Fig. 6.12. The OCF controls rating, credit control, and account balance management. Account balance management is performed by debiting the subscriber's account when interim CCRs are received containing the amount of credits used for resources consumed at the IPTV AS. At the same time the OCF allocates new credits equal to used credits in the previous interim period and sends these to the IPTV AS in a credit control answer (CCA). An elaborate layout of the OCS in the charging framework is illustrated in Fig. 6.9. The source code of the OCS is presented in appendix E.1.6 and released as open source for the purpose of further development and usage in research.

Services		Return					Add
No.	Once off Charge	Duration Charge	Volume Charge	Update	Remove		
1	10	60	0	edit	delete		

Figure 6.11: Service rating information

Services		Return				
No.	Name	Domain	Prepaid Balance	Postpaid Balance	Status	Update
1	alice	mscore.ims	178	1340	1	edit
2	bob	mscore.ims	50	920	1	edit
3	vitalis	mscore.ims	0	0	1	edit
4	gavole	mscore.ims	0	0	1	edit

Figure 6.12: Subscriber account balance information

Offline Charging System

The offline charging system (OFCS) also utilises the Fokus CDP for Diameter interfacing with the IPTV AS. It is implemented in the same way like the online charging system, but it is simplified by elimination of credit control and account balance management

functionality. Subscriber data and billing information is stored in a MySQL database that is shared by the OCS. Charging information generated by the CTF is received across the Rf interface during and after IPTV service usage. The charging data function controls debiting of subscriber accounts in the same manner like with online charging; this is done after rating the service to translate resource usage units to actual credits. Administration of offline charging user accounts is performed using the same interface shown in Fig. 6.12, while Fig. 6.9 illustrates the evaluation layout of the OFCS. The source code of the offline charging system is given in appendix E.1.6 and released as open source. It is programmed in C and runs on Ubuntu Linux.

Volume Charging

Volume charging is implemented using the PCRF and PCEF under the control of CTF in the IPTV AS. The PCRF and PCEF have Diameter modules implemented using the Fokus CDP package to create Rx and Gx interfaces. For evaluation purposes the PCEF counts the number of bytes belonging to a media flow involving an IMS client receiving a stream of packets from a media server. To trigger the charging process the CTF creates an AA-Request message with the IP address of the requesting IMS client and the target media server, including the unique charging ID and sends it to the PCRF. After creating a charging rule in the format shown below the PCRF sends a push RA-Request to the PCEF to create packet filters and initialise IP accounting.

```
init_fw voip src_host_ip dst_host_ip udp subscriberID
```

In the rule `init_fw` is a procedure existing at the PCEF that utilises the Linux IPtables package to create IP accounting chains for each media flow. The flows are identified by the source and destination IP addresses, i.e. the IP address of the client and server. An example of a command to create two-way packet filters at the PCEF is given below; the command is part of the `init_fw` process that is triggered by the PCRF.

```
iptables -A FORWARD -p $UDP -s $src_host_ip/32 -d $dst_host_ip/32 -j $userx  
iptables -A FORWARD -p $UDP -d $src_host_ip/32 -s $dst_host_ip/32 -j $userx
```

The location of the PCEF in the evaluation platform is illustrated in Fig. 6.8; it bridges the IP access and core domains. The IPtables FORWARD chains are hence traversed by media packets.

6.5 Summary

This chapter presented the evaluation framework for service outsourcing and charging in IMS. It comprises a user agent residing in an IMS client and a network agent residing in an IPTV AS. Service outsourcing in IMS utilises a charging platform. The outsourcing framework utilises standard IMS protocols and session procedures provided by the Fokus Open IMS core as well the UCT IMS client and IPTV platform. These are programmed in C, have been tested on Ubuntu Linux and released as open source. The outsourcing and charging framework can be extended and used in IMS and related research. Areas of further development include enforcing service outsourcing on the candidate domain to achieve a fully functional outsourcing scenario that can involve multiple domains. The framework is used to conduct a set of tests aimed at proving the functionality and to evaluate and analyse different outsourcing scenarios.

Chapter 7

Testing and validation

In this chapter a series of tests aimed at establishing the effect of service outsourcing on session charges and revenue for the custodian network are conducted. Moreover, an investigation of the technical effect of introducing outsourcing to general IMS procedures e.g. session setup delay is conducted. The tests further serve as proof of concept of service outsourcing and charging in IMS using IPTV VoD services delivered using the evaluation framework. These tests are performed on the framework discussed in earlier chapters.

Outsourcing results validate the path traversed by SIP registration and session establishment messages, while the performance of the online charging system is monitored and presented. Test results show the range of delays experienced for scenarios ranging from session establishment without addition of charging and outsourcing functionality on the UCT IPTV framework, to session establishment delays for blocked¹ IMS sessions. Finally results showing statistics of call control by the outsourcing framework are presented illustrating the number of accepted, outsourced and rejected calls, and the distribution of the call events when the call waiting periods and durations are changed.

7.1 Proof of Concept Tests

To observe the outsourcing process a trace of the path traversed by SIP register and session establishment messages from the IMS client to the IPTV AS is made. The messages are monitored from sending the first INVITE until a session is successfully

¹sessions that could not be admitted or outsourced

established with the IPTV application server. The aim is to verify that SIP messages traverse the custodian network prior to outsourcing and the candidate network after. Validation of the functioning of the IMS charging system is accomplished by monitoring credit control for session based online charging.

7.1.1 IMS Charging System

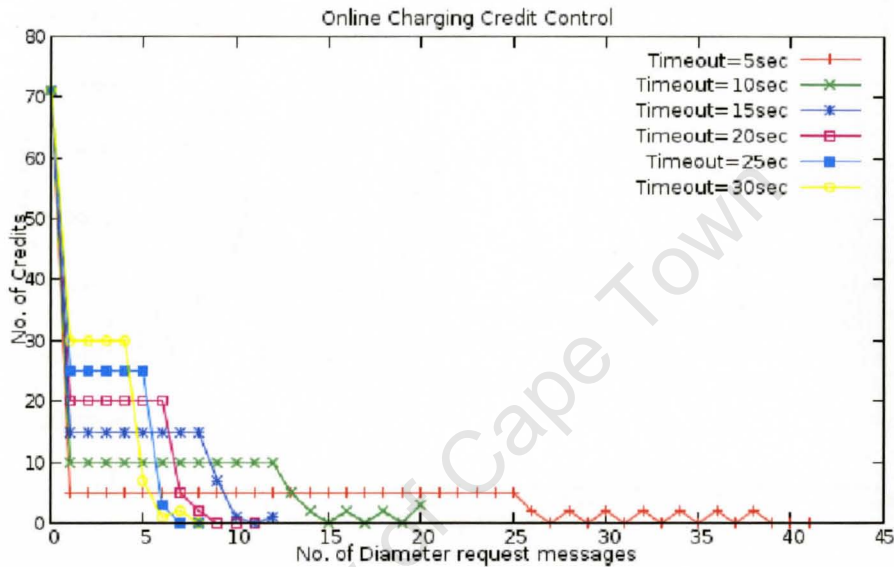


Figure 7.1: Credit control by the online charging system

Upon arrival of an INVITE at the IPTV AS the integrated CTF triggers online charging and sends a CCR to the OCS. The OCS is configured to allocate 10 credits towards a once off service fee, and 60 credits for duration based charging, with each credit translating to 1 second of service usage. The IPTV AS triggers service delivery to the user after receiving a CCA from the OCS with credit allocations. When the IPTV AS sends interim CCR messages to the OCS the reported credit usage is updated to the user profile and new credits are allocated to ensure continued service delivery. Several tests are performed on a user's account that has 200 credits at the start of each cycle.

The tests involve establishing an IPTV VoD session and use different interim periods ranging from 5 to 30 seconds. Each session is terminated by the IPTV AS when credits get exhausted. The results show that when using a 30 seconds interim period sessions are terminated after 8 following sending of the 5th interim report to the OCS. This is

illustrated in Fig. 7.1, where the number of Diameter CCR messages sent to the OCS for different interim report periods are depicted.

In addition to duration charging, volume charging is implemented at the PCEF to measure IPTV media traffic. Figure. 7.2 depicts the bursty nature of streaming video traffic at the media plane.

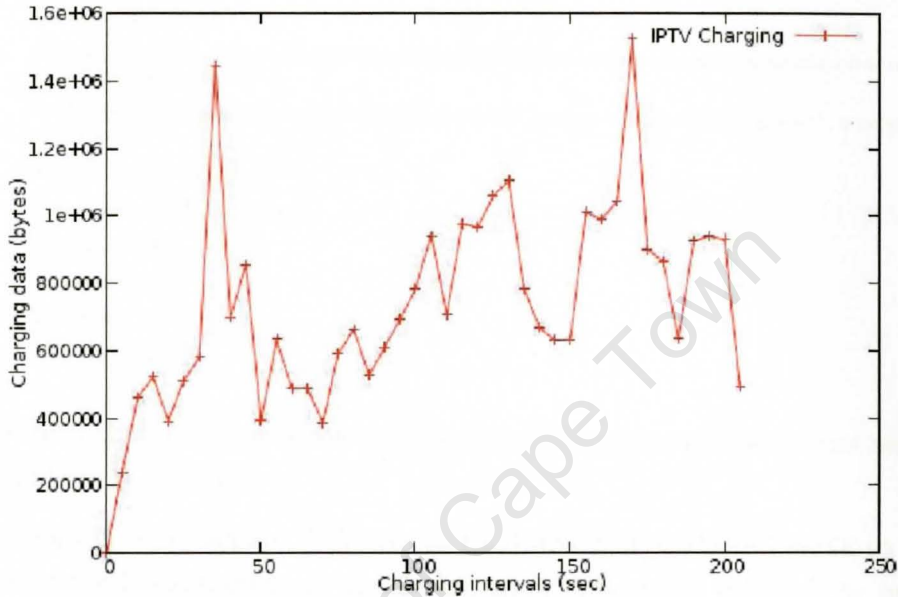


Figure 7.2: IPTV volume charging

The charging results present an experimental proof of the design and performance of the IMS charging framework.

7.1.2 Service Outsourcing: Custodian Network - Setup Phase

A trace of SIP signalling during service outsourcing is conducted by connecting the IMS client to the custodian network and establishing a session. The IP address assigned to the IMS client is 10.10.0.5, and corresponds to the SIP contact address used in registration and session establishment. The outgoing P-CSCF is set to pscf.imscore.ims so that IMS registration requests can be routed via it. Contents of the first registration message are shown below:

```
REGISTER sip:imscore.ims SIP/2.0
```

```
Via: SIP/2.0/UDP 10.10.0.5:5061;rport:branch=z9hG4bK1485306918
Route: <sip:pcscf.imscore.ims:4060;lr>
From: <sip:alice@imscore.ims>;tag=1331743018
To: <sip:alice@imscore.ims>
Call-ID: 225496268
CSeq: 1 REGISTER
Contact: <sip:alice@10.10.0.5:5061;line=048cfa979319bf6>;+sip.instance="<urn:uuid:19cb41d6-64ae-11df-ba0b-2f68965ff8f2>"
Authorization: Digest username="alice@imscore.ims", realm="imscore.ims", nonce=" ", uri="sip:imscore.ims", response=" "
Max-Forwards: 70
User-Agent: UCT IMS Client
Expires: 600000
Supported: path
Supported: gruu
Content-Length: 0
```

It can be seen that as expected the message is routed via pcscf.imscore.ims and the user agent's address is 10.10.0.5.

In an attempt to establish an IPTV session the IMS client sends an INVITE to IPTV AS. Contents of the INVITE message are shown below, where the value of the outsourcing (a=ot:0) attribute in the SDP body is set to zero to indicate to the IPTV AS that the request does not belong to an outsourced session.

```
INVITE sip:channel1@iptv.imscore.ims SIP/2.0
Via: SIP/2.0/UDP 10.10.0.5:5061;rport:branch=z9hG4bK891688932
Route: <sip:pcscf.imscore.ims:4060;lr>
Route: <sip:orig@sccsf.imscore.ims:6060;lr>
From: "Alice" <sip:alice@imscore.ims>;tag=1305647935
To: <sip:channel1@iptv.imscore.ims>
Call-ID: 496913293
CSeq: 20 INVITE
Contact: <sip:alice@10.10.0.5:5061>
Content-Type: application/sdp
Allow: INVITE, ACK, CANCEL, BYE, PRACK, UPDATE, REFER, MESSAGE
Max-Forwards: 70
User-Agent: UCT IMS Client
Subject: IMS Call
Expires: 120
```

```
P-Preferred-Identity: 'Alice' <sip:alice@imscore.ims>
P-Preferred-Service: urn:xxx:3gpp-service.ims.icsi.mmtel
Privacy: none
P-Access-Network-Info: IEEE-802.11a
Require: precondition
Require: sec-agree
Proxy-Require: sec-agree
Supported: 100rel
Content-Length: 341
v=0
o=- 0 0 IN IP4 10.10.0.5
s=IMS Call
c=IN IP4 10.10.0.5
t=0 0
m=audio 26252 RTP/AVP 3 0 101
b=AS:64
a=rtpmap:3 GSM/8000
a=rtpmap:0 PCMU/8000
a=rtpmap:101 telephone-event/8000
a=fmtp:101 0-11
a=curr:qos local none
a=curr:qos remote none
a=des:qos mandatory local sendrecv
a=des:qos mandatory remote sendrecv
a=ct:0
a=ot:0
```

The SIP contents of the INVITE confirm the SIP contact address to be 10.10.0.5 and the message is routed via pscf.imscore.ims, which should be the case of sessions served on the custodian network.

If the session attempt is outsourced the client would receive an outsource command from the IPTV server and de-register with the imscore.ims realm before performing an IP address change to use the imsnet.ims realm's IP-CAN. This is the process of attaching to the candidate network. The de-registration message sent by the client is shown below:

```
REGISTER sip:imscore.ims SIP/2.0
Via: SIP/2.0/UDP 10.10.0.5:5061;rport;branch=z9hG4bK91457114
Route: <sip:pcscf.imscore.ims:4060;lr>
From: <sip:alice@imscore.ims>;tag=1331743018
To: <sip:alice@imscore.ims>
Call-ID: 225496268
CSeq: 3 REGISTER
Contact: <sip:alice@10.10.0.5:5061;line=048cfa979319bf6>;+sip.instance="<urn:uuid:19cb41d6-64ae-11df-
ba0b-2f68965ff8f2>"
Authorization: Digest username="alice@imscore.ims", realm="imscore.ims", nonce=" ", uri="imscore.ims",
response=" "
Max-Forwards: 70
User-Agent: UCT IMS Client
Expires: 0
Supported: path
Supported: gruu
Content-Length: 0
```

This message originates from 10.10.0.5 and is routed via pcscf.imscore.ims on the custodian network.

7.1.3 Service Outsourcing: Candidate Network - Execution Phase

After attaching to the candidate network the IP address assigned to the IMS client is changed to 10.200.0.5. To accomplish outsourcing, the IMS client performs a second registration with the imscore.ims core using pcscf.imsnet.ims. The contents of the registration message are shown below:

```
REGISTER sip:imscore.ims SIP/2.0
Via: SIP/2.0/UDP 10.200.0.5:5061;rport;branch=z9hG4bK1840894505
Route: <sip:pcscf.imsnet.ims:4060;lr>
From: <sip:alice@imscore.ims>;tag=119664517
To: <sip:alice@imscore.ims>
Call-ID: 888533253
CSeq: 1 REGISTER
```

```
Contact: <sip:alice@10.200.0.5:5061;line=048cfa979319bf6>;+sip.instance="<urn:uuid:1a7ff2be-64f6-11df-
ba0b-2f68965ff8f2>"
Authorization: Digest username="alice@imscore.ims", realm="imscore.ims", nonce=" ", uri="sip:imscore.ims",
response=" "
Max-Forwards: 70
User-Agent: UCT IMS Client
Expires: 600000
Supported: path
Supported: gruu
Content-Length: 0
```

It illustrates that the message is sent to imscore.ims core network, but is routed via pscf.imsnet.ims on the candidate network.

After successful registration the client sends a new INVITE to the IPTV AS in an attempt to establish the session. Similar to the second registration the INVITE is routed via the candidate IP-CAN and the candidate network's P-CSCF. The contents of the INVITE message, in which the outsourcing optional attribute is set to 1 (a=ot:1) is shown below. This indicates to the IPTV server that the request belongs to an outsourced session.

```
INVITE sip:channel1@iptv.imscore.ims SIP/2.0
Via: SIP/2.0/UDP 10.200.0.5:5061;rport;branch=z9hG4bK 1862447991
Route: <sip:pscf.imsnet.ims:4060;lr>
Route: <sip:orig@scscf.imscore.ims:6060;lr>
From: "Alice" <sip:alice@imscore.ims>;tag=1549690087
To: <sip:channel1@iptv.imscore.ims>
Call-ID: 555198282
CSeq: 20 INVITE
Contact: <sip:alice@10.200.0.5:5061>
Content-Type: application/sdp
Allow: INVITE, ACK, CANCEL, BYE, PRACK, UPDATE, REFER, MESSAGE
Max-Forwards: 70
User-Agent: UCT IMS Client
Subject: IMS Call
Expires: 120
P-Preferred-Identity: "Alice" <sip:alice@imscore.ims>
P-Preferred-Service: urn:xxx:3gpp-service.ims.icsi.mmstel
Privacy: none
```

```
P-Access-Network-Info: IEEE-802.11a
Require: precondition
Require: sec-agree
Proxy-Require: sec-agree
Supported: 100rel
Content-Length: 341
v=0
o=- 0 0 IN IP4 10.200.0.5
s=IMS Call
c=IN IP4 10.200.0.5
t=0 0
m=audio 26252 RTP/AVP 3 0 101
b=AS:64
a=rtpmap:3 GSM/8000
a=rtpmap:0 PCMU/8000
a=rtpmap:101 telephone-event/8000
a=fmtp:101 0-11
a=curr:qos local none
a=curr:qos remote none
a=des:qos mandatory local sendrecv
a=des:qos mandatory remote sendrecv
a=ct:0
a=ot:1
```

7.1.4 Outsourcing Network Agent Messages

The IPTV server responds to INVITE requests from clients with messages containing different SIP codes and content. A successful INVITE request triggers a 200OK response that contained a URL for the requested media. An example of the successful response for a session that is admitted to the custodian network is shown below.

```
SIP/2.0 200 OK
Via: SIP/2.0/UDP 10.20.0.3:6060;branch=z9hG4bK5668.d0ae7b71.0
Via: SIP/2.0/UDP 10.20.0.3:4060;branch=z9hG4bK5668.8e4d4e91.0
Via: SIP/2.0/UDP 10.10.0.5:5062;rport=5062;branch=z9hG4bK1000321492
Record-Route: <sip:mo@scscf.imscore.ims:6060;lr>
```

```
Record-Route: <sip:mo@pcscf.imscore.ims:4060;lr>
From: "Bob" <sip:bob@imscore.ims>;tag=321145986
To: <sip:channel3@iptv.imscore.ims>;tag=997407924
Call-ID: 455225860
CSeq: 20 INVITE
Contact: <sip:channel3@10.20.0.3:8010>
Content-Type: message/external-body; access-type="URL"; expiration="Sat, 01 January 2011 09:00:00
  GMT"; URL="rtsp://tv.imscore.ims:5554/channel3"
User-Agent: eXosip/3.0.3
Content-Length: 0
```

When a session is outsourced the IPTV server sends a message with the SIP code 491 to indicate a pending request. This message commands the IMS client to perform a handover to the indicated candidate network; the value of OCN indicates the name of the candidate network to be used. An example of the SIP 491 message is shown below.

```
SIP/2.0 491 Request Pending
Via: SIP/2.0/UDP 10.20.0.3:6060;branch=z9hG4bK6e65.541e26.0
Via: SIP/2.0/UDP 10.20.0.3:4060;branch=z9hG4bK6e65.ab994615.0
Via: SIP/2.0/UDP 10.10.0.5:5062;rport=5062;branch=z9hG4bK1157330567
From: "Bob" <sip:bob@imscore.ims>;tag=158807489
To: <sip:channel3@iptv.imscore.ims>;tag=1443517842
Call-ID: 138592556
CSeq: 20 INVITE
Content-Type: message/external-body; access-type="URL"; expiration="Sat, 01 January 2011 09:00:00
  GMT"; OCN="imsnet.ims"
User-Agent: eXosip/3.0.3
Content-Length: 0
```

When the IPTV server receives an INVITE for an outsourced session it responds with a 200OK containing the URL of requested media. This response is routed via the candidate network's P-CSCF used for the request. Moreover, the SIP contact address for the IMS client corresponds to the new IP address assigned on the candidate IP-CAN. An example of a response to an outsourced request is given below. Although the URL of the requested media is tv.imscore.ims, the media traffic for the outsourced session is routed via the candidate IP-CAN.

```
SIP/2.0 200 OK
Via: SIP/2.0/UDP 10.20.0.3:6060;branch=z9hG4bKfac5.07fe0ba2.0
Via: SIP/2.0/UDP 10.20.0.3:4060;branch=z9hG4bKfac5.3cbf95f6.0
Via: SIP/2.0/UDP 10.200.0.5:5062;rport=5062;branch=z9hG4bK808496183
Record-Route: <sip:mo@scscf.imscore.ims:6060;lr>
Record-Route: <sip:mo@pcscf.imsnet.ims:4060;lr>
From: "Bob" <sip:bob@imscore.ims>;tag=106910731
To: <sip:channel3@iptv.imscore.ims>;tag=609495566
Call-ID: 1835803884
CSeq: 20 INVITE
Contact: <sip:channel3@10.20.0.3:8010>
Content-Type: message/external-body; access-type="URL"; expiration="Sat, 01 January 2011 09:00:00
  GMT"; URL="rtsp://tv.imscore.ims:5554/channel3"
User-Agent: eXosip/3.0.3
Content-Length: 0
```

If a session request is blocked the IPTV server responds to the client with a message with SIP code 480 to indicate unavailability of the requested service. An example of the 480 response for a blocked session is shown below.

```
SIP/2.0 480 Temporarily not available
Via: SIP/2.0/UDP 10.20.0.3:6060;branch=z9hG4bKc894.36797836.0
Via: SIP/2.0/UDP 10.20.0.3:4060;branch=z9hG4bKc894.29334374.0
Via: SIP/2.0/UDP 10.10.0.5:5061;rport=5061;branch=z9hG4bK2021124458
From: "Alice" <sip:alice@imscore.ims>;tag=685204070
To: <sip:channel1@iptv.imscore.ims>;tag=1997642837
Call-ID: 1578968532
CSeq: 20 INVITE
User-Agent: eXosip/3.0.3
Content-Length: 0
```

Different server messages are interpreted by the client to trigger pre-defined actions e.g., outsourcing or a re-try of session establishment for a blocked session.

7.2 Session Establishment Delay

In the outsourcing setup phase tests are performed to establish the delay incurred for session establishment of IPTV services. These involve scenarios when charging is deactivated, when charging is activated and when service outsourcing is activated. The delay from the moment the IMS client sends the first INVITE to the IPTV server until the reception of a successful response containing a 200OK is recorded .

In first scenario tests are performed to compare the session setup delay for an IMS system on which charging is not implemented and a system where charging is invoked at the service level only.

7.2.1 Service Level Charging

Upon receiving an INVITE the IPTV AS sends a Diameter message to the OCS in the billing domain for credit control and session authorisation. This is followed by sending a response with a 200OK in a message containing a media URL for the requested channel, hence marking the completion of the test cycle.

A total of 50 sessions are established and setup delays for each recorded. It should be noted that delays for different test instances are independent of other sessions. The observed delay pattern is illustrated in Fig. 7.3. This scenario is used as a reference for comparison with subsequent scenarios.

From the graph it is evident that session setup delays without charging and with service level charging are comparatively equal. The predominant delay range is 800 to 1300 mill seconds. This means service level charging does not introduce additional delay to session establishment. The variation in delays experienced by different instances is a result of network conditions in an Ethernet environment that is governed by network contention for media resources.

7.2.2 Policy and Charging Control on the Media Plane

In this scenario the delay incurred for session establishment with volume charging is recorded. Volume charging is performed using PCC procedures as discussed earlier. Figure 7.4 illustrates the recorded delay pattern.

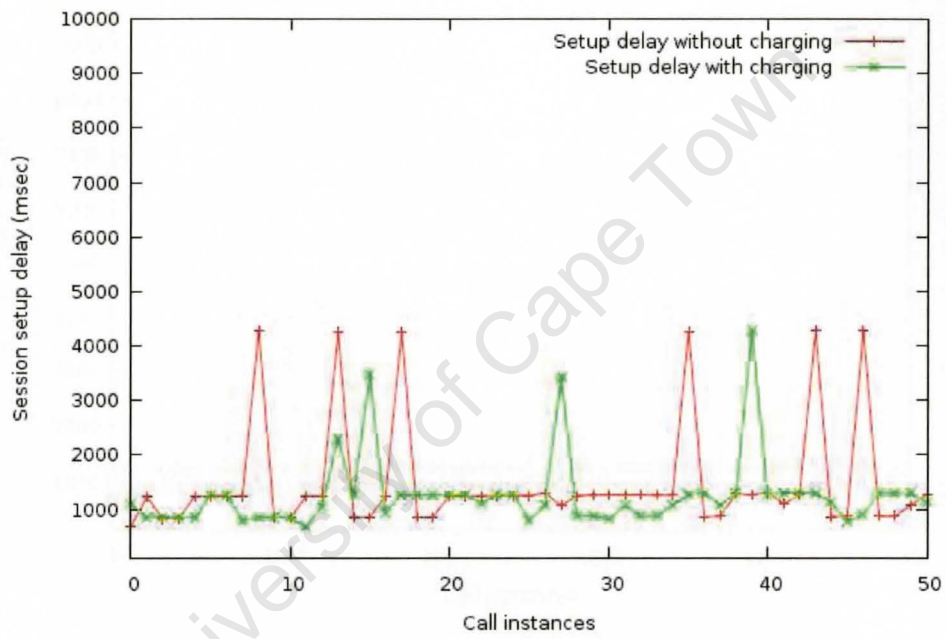


Figure 7.3: IPTV Session setup delay

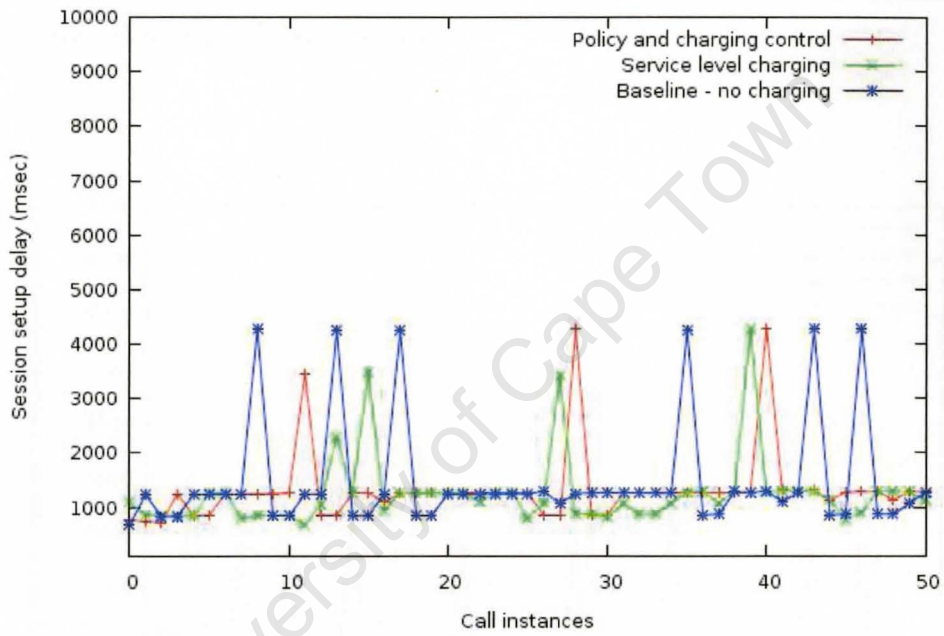


Figure 7.4: Session setup delay with PCC procedures

When compared to the baseline results, it is evident that PCC procedures do not introduce noticeable delay to session establishment.

7.2.3 Service Outsourcing - Sessions Admitted on Custodian Network

In scenarios involving service outsourcing tests are conducted to determine the delay incurred for sessions that are processed by the outsourcing network agent and eventually admitted to the custodian network. Figure 7.5 illustrates the session establishment delay incurred for sessions involving four users.

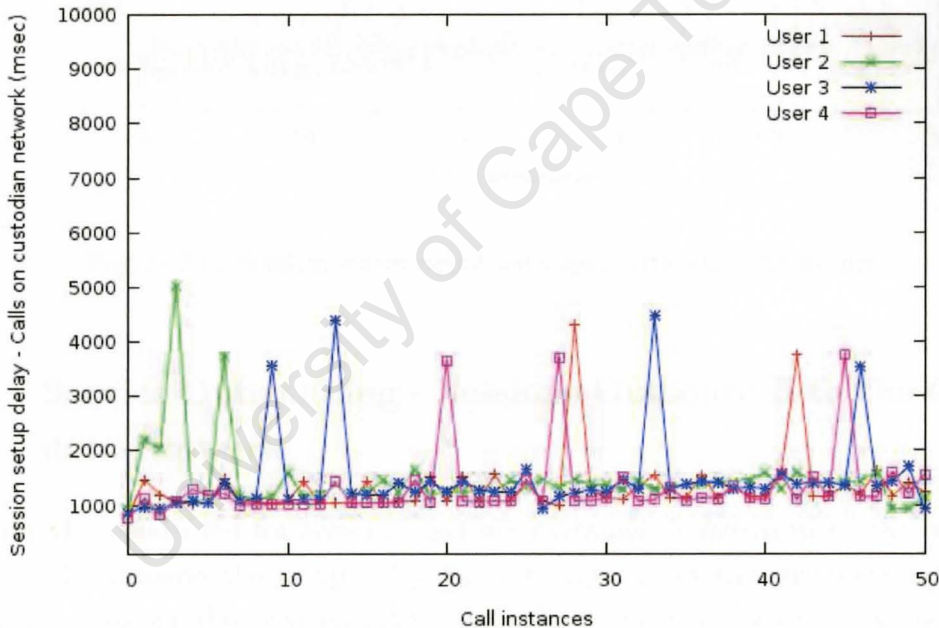


Figure 7.5: Setup delays for sessions admitted on the custodian network

The delays range between 1000 to 1300 msec, which is higher than the baseline results presented above. This means that service outsourcing mechanisms introduce some delay to IPTV session establishment. However, the additional delay would not inconvenience

users of IPTV services since they are incurred at session initiation time. A comparison between the delay incurred due to outsourcing mechanisms and the previous scenarios (without outsourcing) is illustrated in Fig. 7.6.

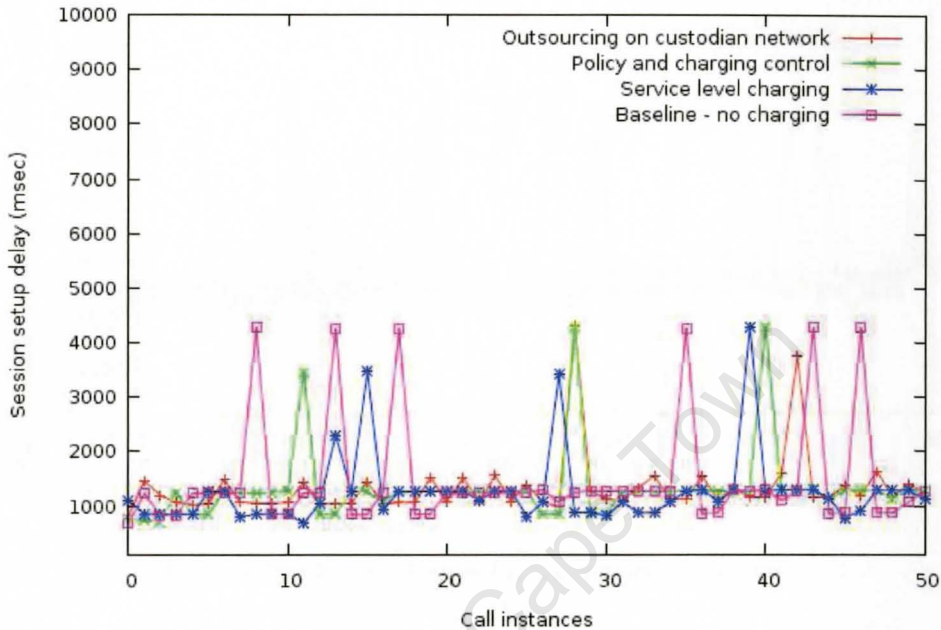


Figure 7.6: Session setup delay with and without outsourcing

7.2.4 Service Outsourcing - Sessions Outsourced to the Candidate Network

Tests are also conducted for sessions that are eventually outsourced to the candidate network. To measure the setup delay the full capacity of the custodian network is allocated to sessions that remain active throughout the test duration. All additional session establishment attempts are subject to outsourcing to the candidate network. The delays experienced by outsourced sessions for 4 users are shown in Fig. 7.7. It is clear from the graph that outsourced sessions incur a higher setup delay ranging between 2000 and 2300 msec.

A comparison between delays of outsourced sessions and previous scenarios is given in Fig. 7.8. The graph highlights the difference and indicates that outsourcing resulted in

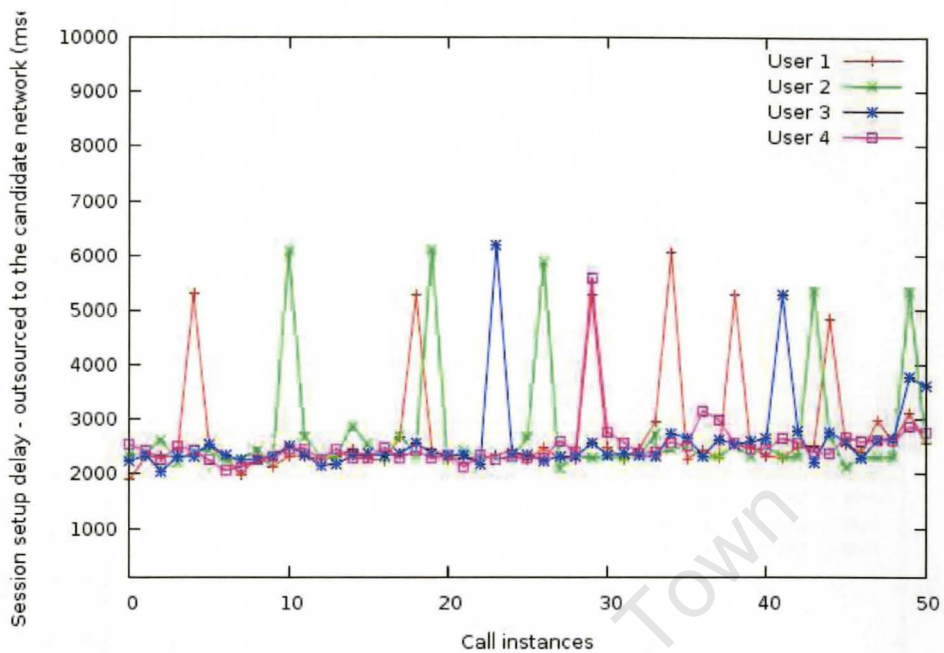


Figure 7.7: Setup delays for outsourced sessions

double the amount of delay for session establishment. Although an additional 1000 msec delay would be noticeable by users it is considered insignificant when compared to the waiting times for blocked sessions.

7.2.5 Service Outsourcing - A General Scenario

Figure 7.9 and Fig. 7.10 show the delays recorded for session establishment attempts made by users over a 5 minutes period. The graphs represent two scenarios that differ by the duration of calls admitted on the custodian and candidate networks. Full details of test scenarios are presented below. The graph in Fig. 7.9 applies to scenario 1a (refer to section 7.3), while Fig. 7.10 applies to scenario 2a (refer to section 7.3).

Call waiting times for blocked attempts contributed to the long setup delays, which vary between 5 and 25 seconds when the call duration is 30 seconds and between 10 and 40 seconds for a 90 seconds call duration. These long call waiting times would frustrate users and may contribute to network churn due to poor service experience. Service outsourcing reduces the call waiting times to improve user experience.

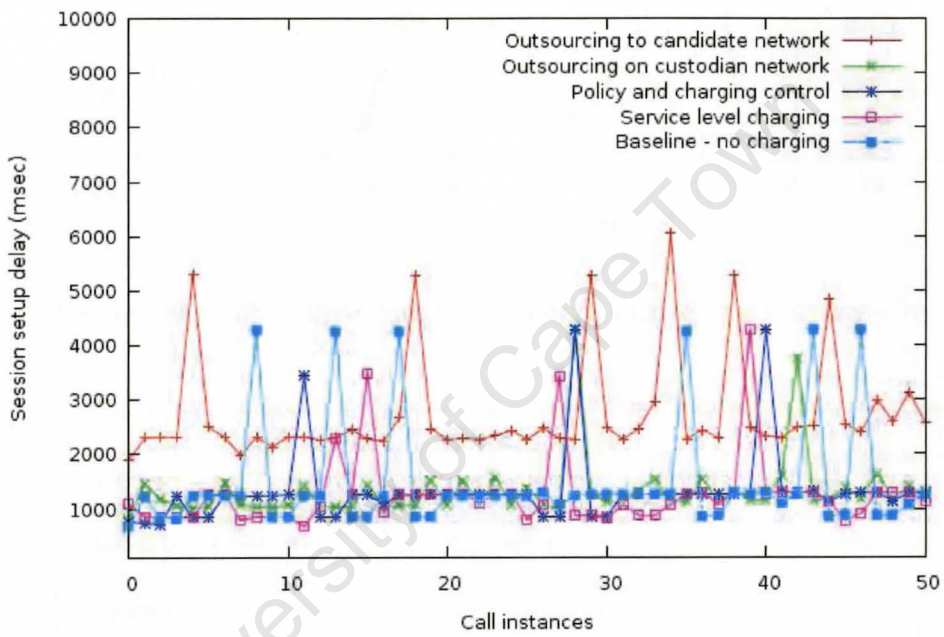


Figure 7.8: Delays for outsourced and admitted sessions

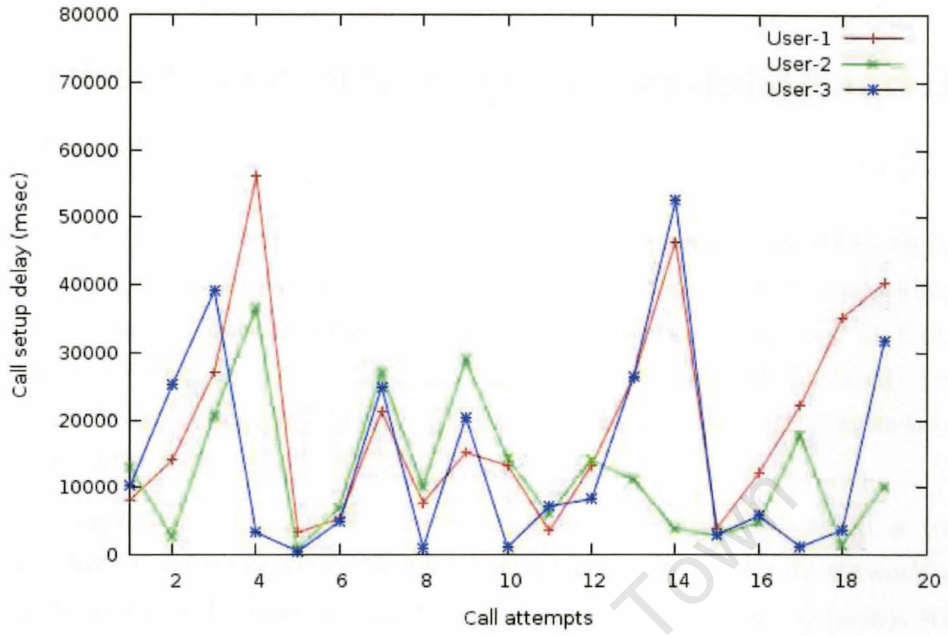


Figure 7.9: Call setup delays for a comprehensive case - scenario 1a

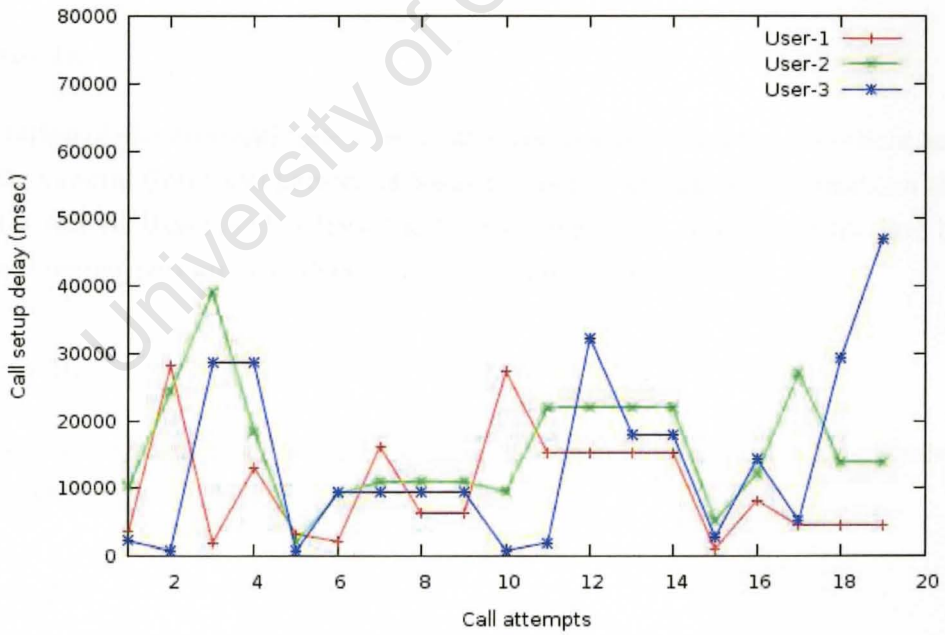


Figure 7.10: Call setup delays for a comprehensive case - scenario 2a

7.3 Distribution of Custodian, Candidate and Blocked Calls

The performance of the outsourcing framework is investigated to establish the distribution of IPTV sessions on custodian and candidate networks, as well as blocking of calls (sessions). A total of 7 scenarios are used to study the outsourcing trend and investigate the effect of changing parameters like call duration, re-try intervals for blocked calls and inter-call intervals for successful calls on the performance of the custodian network. The tests are setup as follows:

For all scenarios the capacity of the custodian network is set to admit a total of 7 sessions. Three more sessions can be accommodated on the candidate network to create a virtual capacity of 10 sessions. 12 IMS clients compete for this capacity, with each client attempting to establish one session. A number of session establishment attempts are to be blocked if all clients attempt to establish sessions. Each test is conducted over a period of 5 minutes.

Scenario 1a

The duration of all successful calls is set to 30 seconds after which the IMS client activates a waiting interval (inter-call period) of 3 seconds before attempting to establish the next call. If a call establishment attempt is blocked the client activates a blocked holding interval (holding period) of 5 seconds before making a new attempt.

Scenario 1b

In this scenario the inter-call period is 3 seconds, the holding period is 15 seconds, while the call duration is set to 30 seconds.

Scenario 2a

The call duration is set to 90 seconds, the inter-call period is 3 seconds and the holding period is 5 seconds.

Scenario 2b

In this scenario the call duration is set to 90 seconds, the inter-call period is 3 seconds and the holding period is 15 seconds.

Scenario 3a

The call duration is set to 60 seconds, the inter-call period is 3 seconds and the holding period is 5 seconds.

Scenario 3b

The call duration is 60 seconds, the inter-call period is 15 seconds and the holding period is 15 seconds as well.

Scenario 4

In scenario 4 the call duration is set to 60 seconds, the inter-call period is 15 seconds and the holding period is 5 seconds.

Test results are collected for each scenario and analysed in three categories i.e., call treatment at arrival time, the utilisation of the network capacity and the expected custodian operator income at call arrival time. Call treatment at arrival time statistics are analysed in terms of sessions that are admitted to the custodian network, sessions outsourced to the candidate network and blocked sessions.

The number of active sessions on the custodian and candidate networks is monitored in real-time and used to represent the utilisation of the physical capacity of the custodian network as well as the capacity leased from the candidate network. The virtual capacity utilisation by sessions admitted on custodian and candidate network is used to analyse and illustrate potential financial implications of outsourcing. Finally, the benefits of service outsourcing to the custodian operator are presented in terms of the income retained from charges levied for sessions at call arrival time.

7.3.1 Treatment at Call Arrival - Admitted, Outsourced and Rejected Sessions

Scenario 1a

The admission, outsourcing and rejection pattern for scenario 1a is shown in Fig. 7.11.

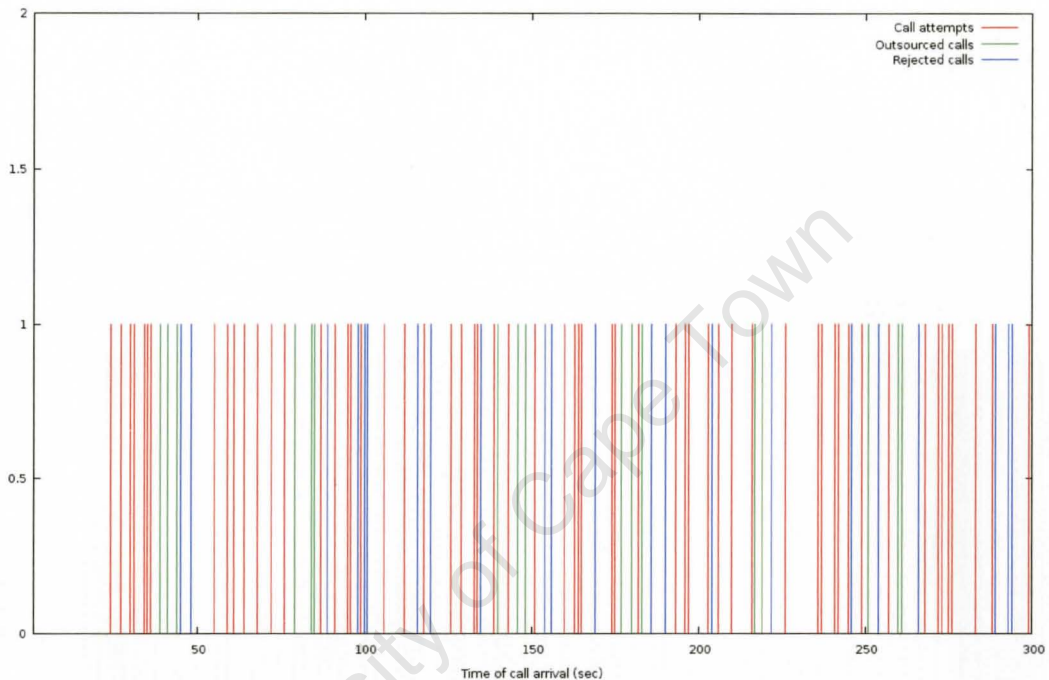


Figure 7.11: Call arrival: Admitted, outsourced and blocked sessions - scenario 1a

The graph shows a pattern depicting when adequate capacity exists on the custodian network to admit arriving calls. The leased capacity is used to outsource calls arriving after the physical capacity on the custodian network is utilised, whereas call blocking occurs after the physical and leased (virtual capacity) are exhausted.

It is evident that the treatment received by arriving calls is fairly distributed across the test period. The occurrence of call blocking instances is also distributed over the test period. This can be attributed to the 30 seconds call duration, which causes calls to terminate early; thus availing capacity for previously blocked calls to be accepted.

Scenario 1b

The call handling pattern for scenario 1b is shown in Fig. 7.12. The graph shows fewer blocked (rejected) calls during the early phase of the test period. Call blocking occurs mainly in the middle and late phases. Generally there are fewer blocked calls when compared to scenario 1a.

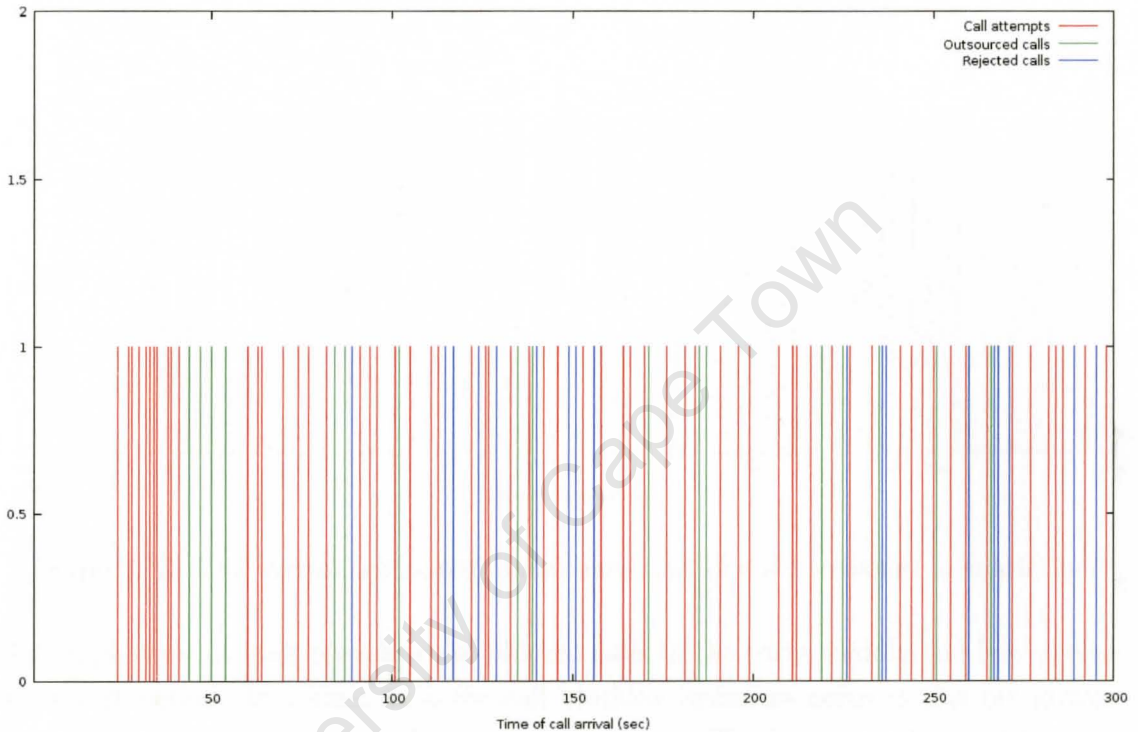


Figure 7.12: Call arrival: admitted, outsourced and blocked sessions - scenario 1b

The occurrence of fewer blocked calls is attributed to the use of a longer holding period of 15 seconds when used in conjunction with a 30 seconds call duration, which means that an IMS client encountering a blocked session may only make 1 re-establishment attempt before an active session is terminated.

Scenario 2a

With respect to scenarios 1a and 1b, in scenario 2a the inter-call and holding periods are shorter in comparison to the duration of active calls. Figure 7.13 shows the call

distribution trend.

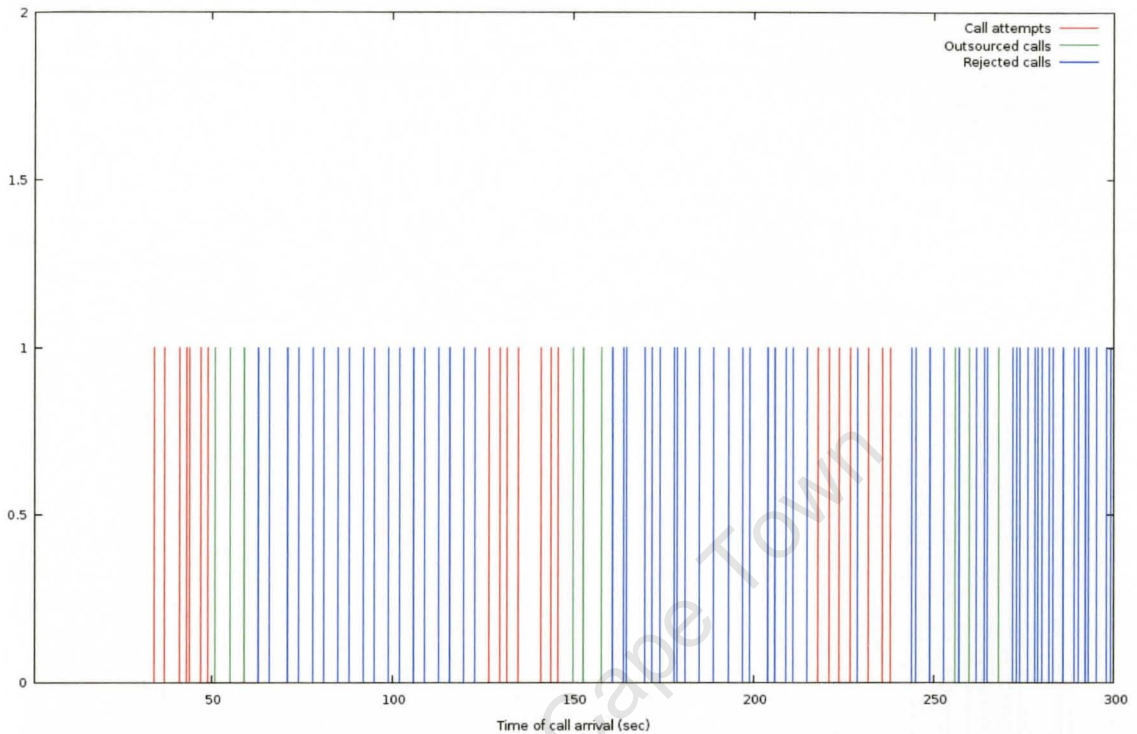


Figure 7.13: Call arrival: admitted, outsourced and blocked sessions - scenario 2a

The graph shows a high prevalence of blocked calls in the early, middle and late phases of the test period. It is clear that the call blocking instances occur in distinct groups that feature after admission and outsourcing phases. The high prevalence of blocked calls can be attributed to IMS clients making multiple new attempts to establish calls while admitted and outsourced sessions are still active.

Scenario 2b

This scenario differs from scenario 2a by the length of the holding period, which is three times longer. The effect of the new value is evident in Fig. 7.14, in which the early phase of the test period has arriving calls admitted and outsourced before call blocking occurs. The clients of blocked calls make at least four more attempts at call establishment before capacity becomes available to accommodate the calls. The middle phase shows a second call admission and outsourcing phase, but it is later dominated by call blocking. The

late phase shows a mixture of the three call treatment types; however, blocked calls are prevalent.

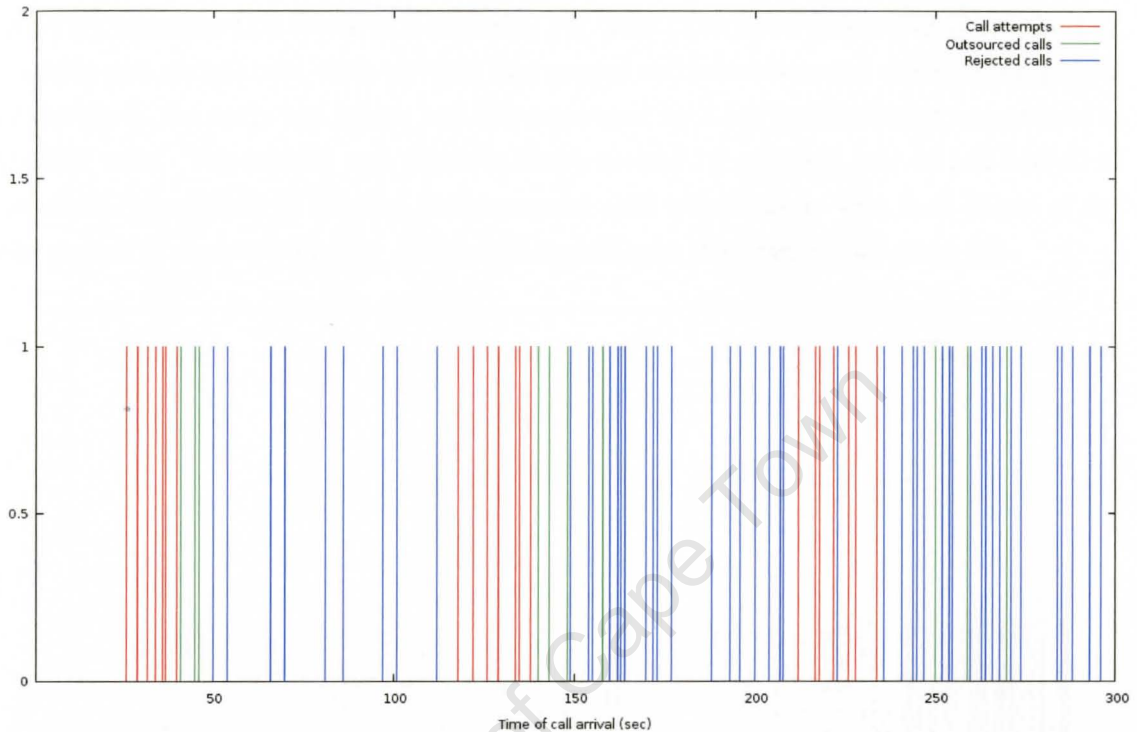


Figure 7.14: Call arrival: admitted, outsourced and blocked sessions - scenario 2b

The pattern seen in the graph highlights the impact of the call holding period on the prevalence of blocked calls as congestion persists. When comparing trends observed in test scenarios 1a and 1b with scenarios 2a and 2b, it is seen that relatively longer holding periods result in overall fewer blocked calls. This can be compared to a real life situation where impatient users of blocked calls make new attempts frequently, while patient users would wait longer before making new attempts.

It is also noted that call duration has a major influence on the prevalence of blocked calls. This is evident from a comparison between scenario 1b and scenario 2b. There is a higher blocked call prevalence in the scenario, which can be related to a real life scenario where new session establishment attempts from users would be blocked as a result of long active sessions occupying the network capacity.

Scenario 3a

The settings of scenario 3a are intermediate between scenarios 1a and 2a. The call handling trend for this scenario is shown in Fig. 7.15. The early phase is characterised by a steady call arrival rate, with the first and second call admission and outsourcing phases occurring in the early test phase and are separated by a fairly distributed occurrence of blocked calls. The middle test phase is characterised by a slower rate of call arrival of which the prevalence of blocked and accepted calls is balanced. The final phase of the test period is characterised by a high call arrival rate that eventually eases off.

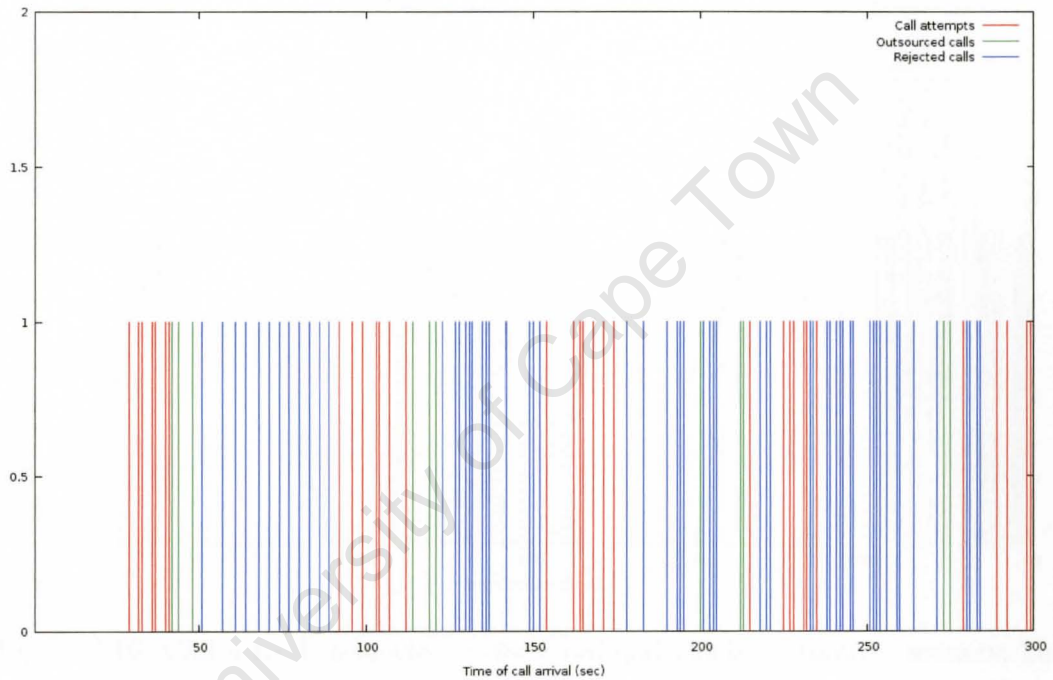


Figure 7.15: Call arrival: admitted, outsourced and blocked sessions - scenario 3a

When compared to scenario 1a a higher call blocking prevalence is seen. However, a lower call blocking prevalence is observed as compared to scenario 2a. This emphasises the relationship between the number of blocked calls and the overall duration of active calls in a given period.

Scenario 3b

The main difference between this scenario and the previous discussed scenarios is the equal length of the inter-call and the holding periods, which are set to 15 seconds. The call duration is 60 seconds implying that blocked IMS clients can make up to 4 new call establishment attempts before network capacity is freed. The call handling trend for this scenario is shown in Fig. 7.16.

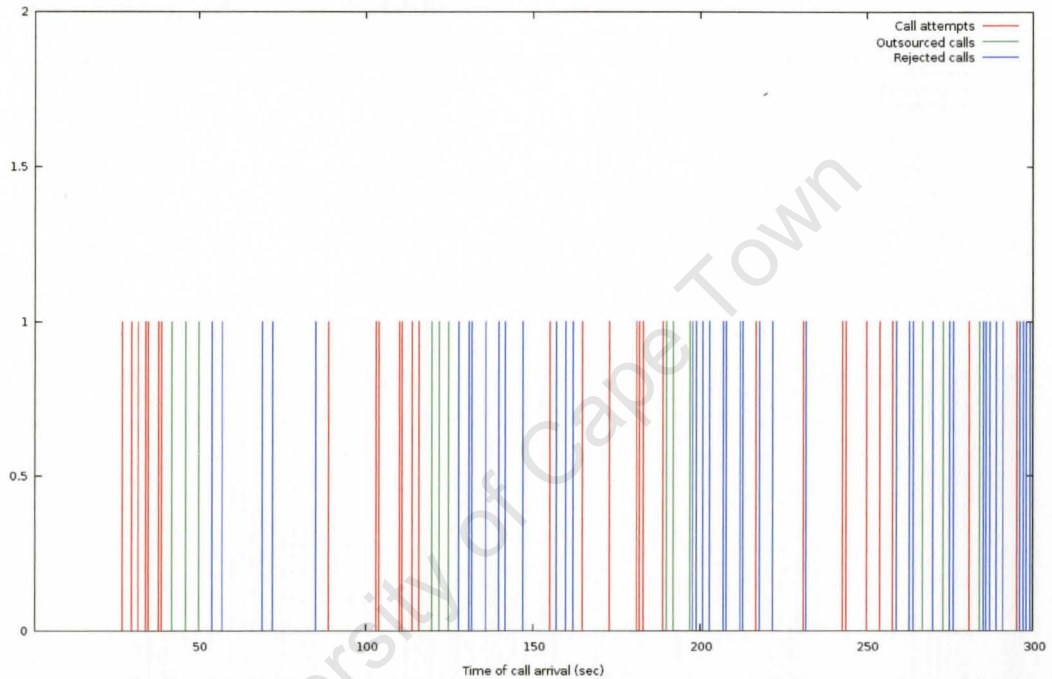


Figure 7.16: Call arrival: admitted, outsourced and blocked sessions - scenario 3b

From the graph it is observed that the early phase of the test period is dominated by call admission to the custodian network and outsourcing of calls to the candidate network, with a few call blocking instances recorded. In the middle phase of the test there is an increase in the call blocking prevalence, which can be attributed to a higher call arrival rate. The beginning and end of the final phase is characterised by a high call arrival rate that is accompanied by a higher call blocking prevalence. When compared to scenario 3a there are more silent periods (without arriving calls). This can be linked to the longer inter-call and holding periods that also result in fewer blocked calls.

It is concluded that longer inter-call and holding periods lead to fewer instances of call

blocking, and under utilisation of network capacity since the demand for resources is erratic and does not coincide with periods of resource abundance.

Scenario 4

Scenario 4 is characterised by an inter-call period that is longer than the holding period. However, the call duration is the same as in scenario 3a and 3b. Figure 7.17 illustrates the call handling pattern for scenario 4.

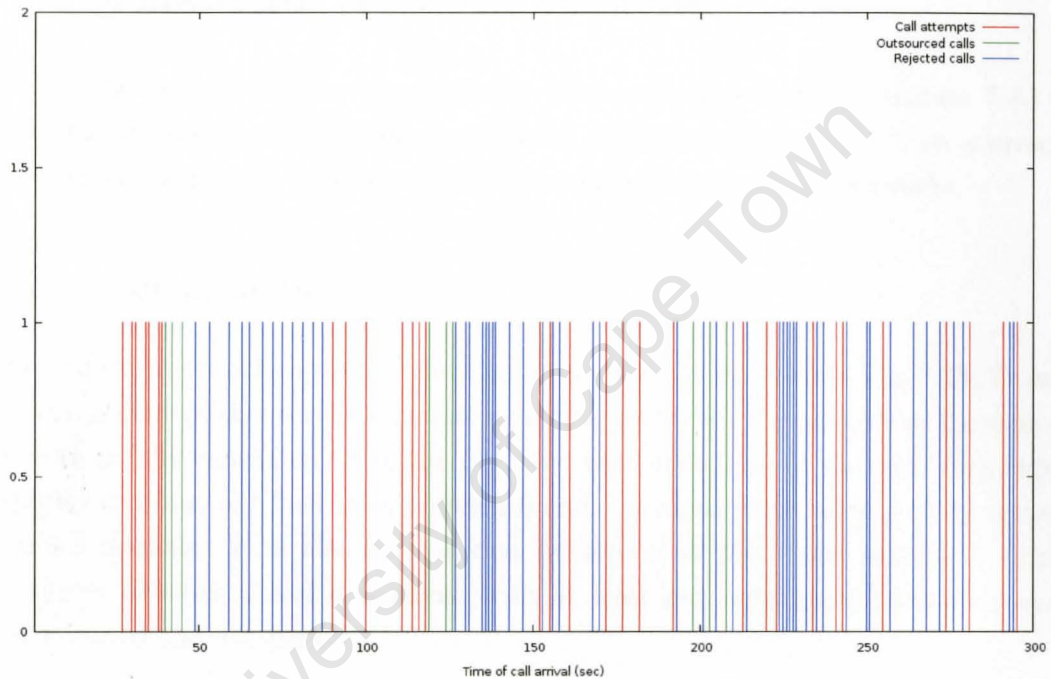


Figure 7.17: Call arrival: admitted, outsourced and blocked sessions - scenario 4

The early phase of the test is characterised by 2 phases of admitted and outsourced calls that are separated by a number of blocked calls and spaced silent periods, while the middle phase of the test period is characterised by a higher prevalence of blocked calls. The blocked call instances are interleaved with admitted and outsourced instances. A similar pattern is observed in the final phase; however, there are distinct bursts of blocked calls before the call arrival rate slows down.

A comparison between scenario 3a and scenario 4 shows a higher call acceptance rate for the latter. This is attributed to the longer inter-call duration that slows down the

call arrival rate. Consequently there is a better interleaving of call treatment in the latter scenario. This can be related to fair sharing of resources by users when calls from blocked users get admitted to the network sooner.

7.4 Network Capacity Utilisation - Custodian and Candidate Network Sessions

During the test periods for the scenarios discussed above (refer to section 7.3.1) the real-time utilisation of the physical and leased capacities are recorded. This is presented in terms of the number of active calls on custodian and candidate networks.

7.4.1 Scenario 1a

The real-time network capacity utilisation for scenario 1a is shown in Fig. 7.18. From the graph it is evident that service outsourcing is triggered when the capacity of the custodian network is fully consumed. It is also observed that utilisation of the custodian network capacity continuously fluctuates between 6 and 7 sessions, with some periods recording 4 and 5 sessions. It is also noticed that utilisation of the leased candidate capacity fluctuates between 2 and 3 sessions, with at least one outsourced session remaining active during the test period.

There are times when the physical capacity on the custodian network is underutilised, while outsourced sessions remain active on the candidate network. During these periods the custodian network loses potential income that would be earned if outsourced sessions had been served locally. The lost income is proportional to the incurred outsource cost margin (OCM) for affected sessions. The extent to which the financial implications of outsourcing are dependent on call duration, inter-call periods and holding periods is discussed later in this chapter.

7.4.2 Scenario 1b

The utilisation of the physical and leased capacities for scenario 1b is shown in Fig. 7.19. As seen from the graph, service outsourcing is triggered in the early phase of the test

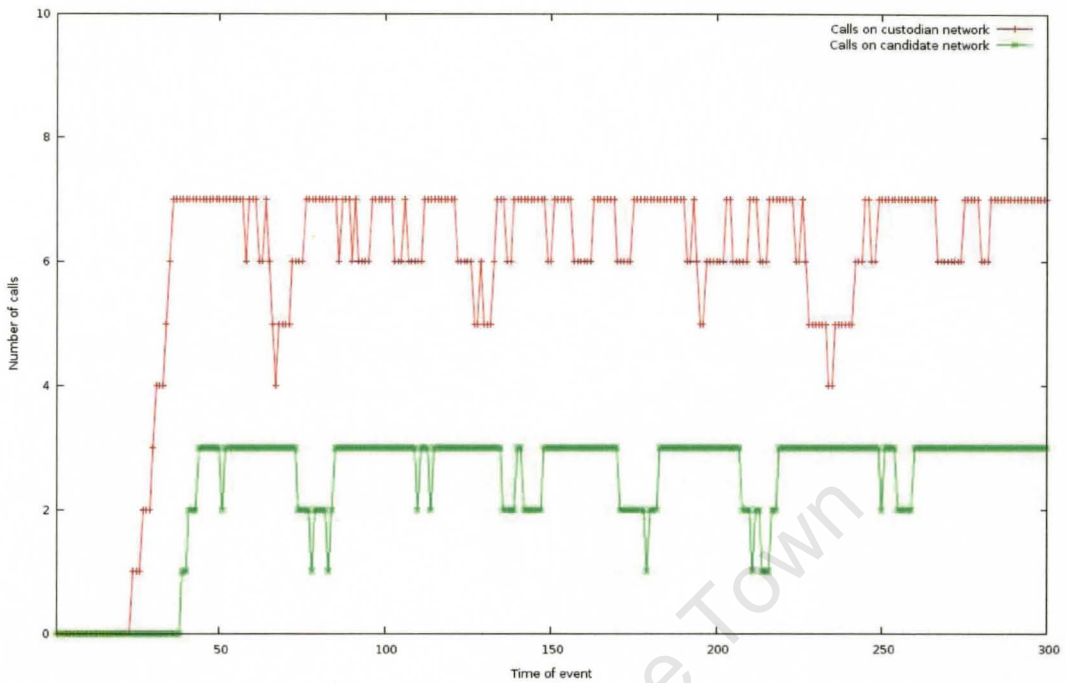


Figure 7.18: The number of custodian and candidate network calls - scenario 1a

period. It is noticed that the physical capacity utilisation fluctuates between 6 and 7 sessions. When compared to scenario 1a, the utilisation of the leased capacity also fluctuates. However, the utilisation of the leased capacity is more conservative as highlighted by an instance between 150 and 200 seconds when the leased capacity is unused.

In general the custodian network benefits from outsourcing but also loses potential income due to outsourcing, as a result of the fluctuating utilisation of the physical capacity.

7.4.3 Scenario 2a

The utilisation of the physical and leased capacity is illustrated in Fig. 7.20. The graph shows that outsourcing is triggered in the early phase of the test period when the physical capacity of the custodian network is exhausted. During the early phase the physical and leased capacities are fully utilised. In general, except for a few instances that last for a total period of less than 20 seconds, the physical capacity is fully utilised from about the 50th second until the end of the test period. The same trend is depicted for the leased capacity utilisation.

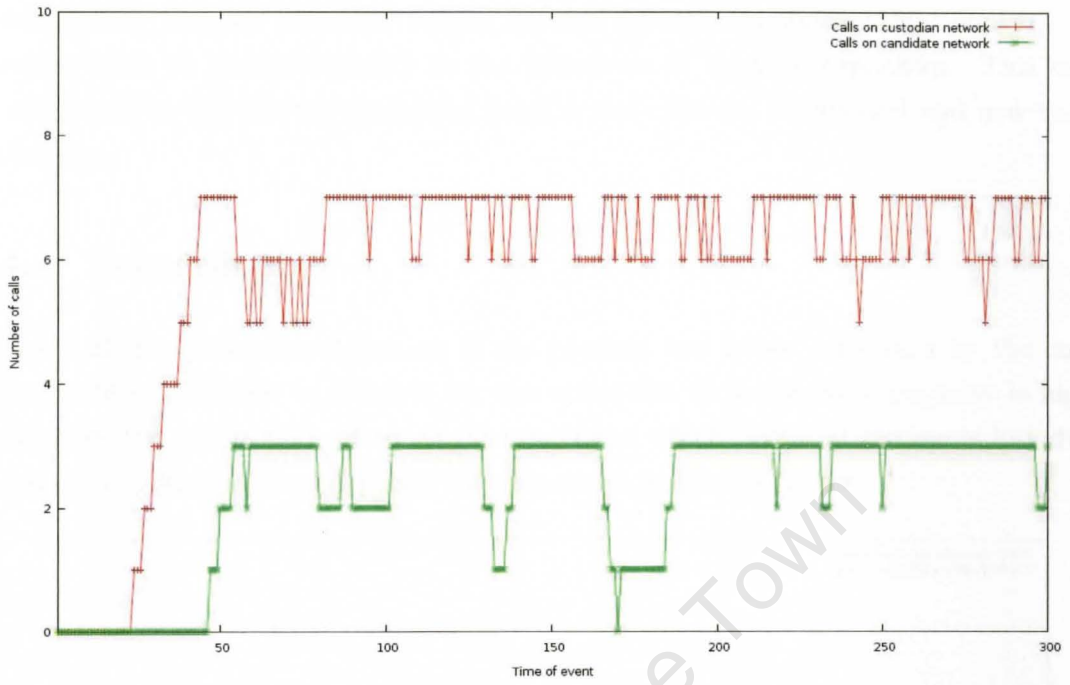


Figure 7.19: The number of custodian and candidate network calls - scenario 1b

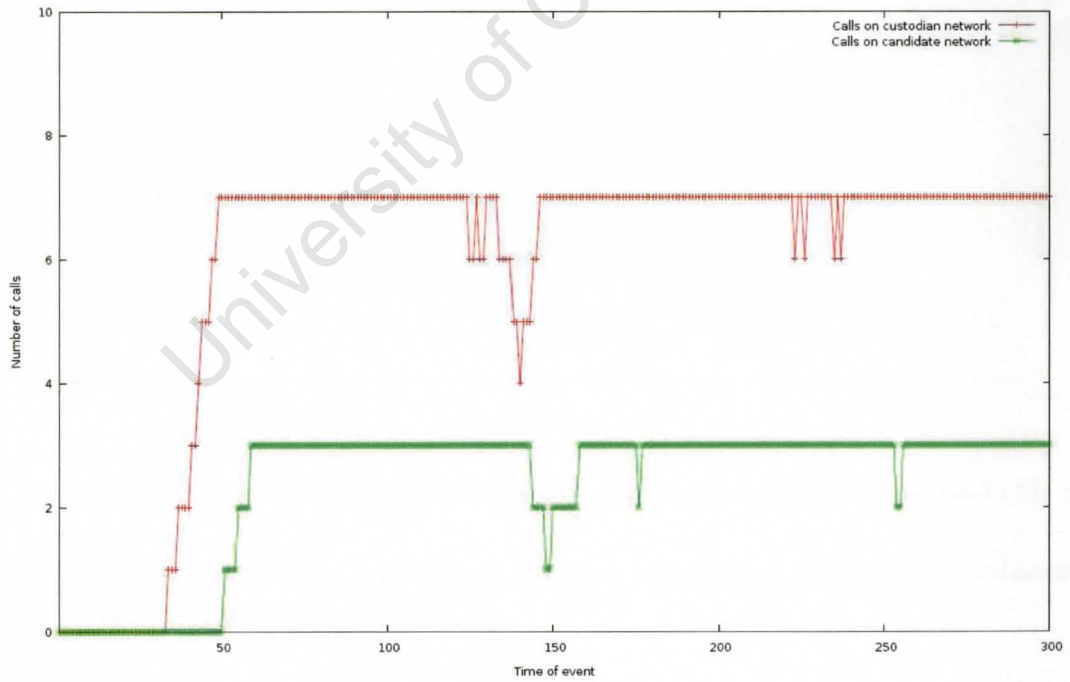


Figure 7.20: The number of custodian and candidate network calls - scenario 2a

A comparison between scenarios 1a and 1b and scenario 2a shows that a longer call duration leads to greater stability in the utilisation of network capacities. This can be attributed to the (fewer) number of times active calls are terminated and new calls established.

7.4.4 Scenario 2b

Figure 7.21 illustrates the utilisation of the physical and leased capacities by the custodian network. Similar to scenario 2a, the utilisation of the physical capacity is high throughout the test period. However, the extent to which potential income is lost due to failure to utilise physical capacity by outsourced sessions is higher.

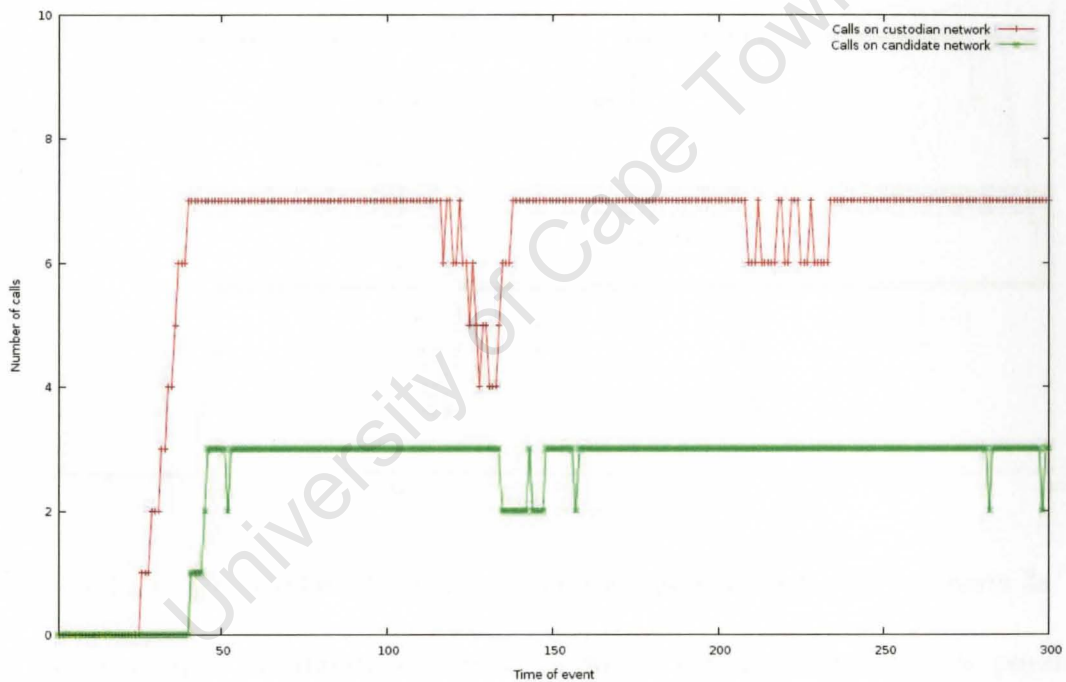


Figure 7.21: The number of custodian and candidate network calls - scenario 2b

It is evident that the holding period does not have a significant effect on the utilisation of the physical and leased capacities.

7.4.5 Scenario 3a

Network capacity utilisation for this scenario is shown in Fig. 7.22. There is more stability in the physical and leased capacity allocation when compared to scenarios 1a and 1b. However, there is a higher degree of fluctuating capacity utilisation when compared to scenarios 2a and 2b. Fluctuating physical capacity utilisation ranges from 3 to 7 sessions; this happens when outsourced sessions are active on the candidate network resulting in higher lost potential income during the period.

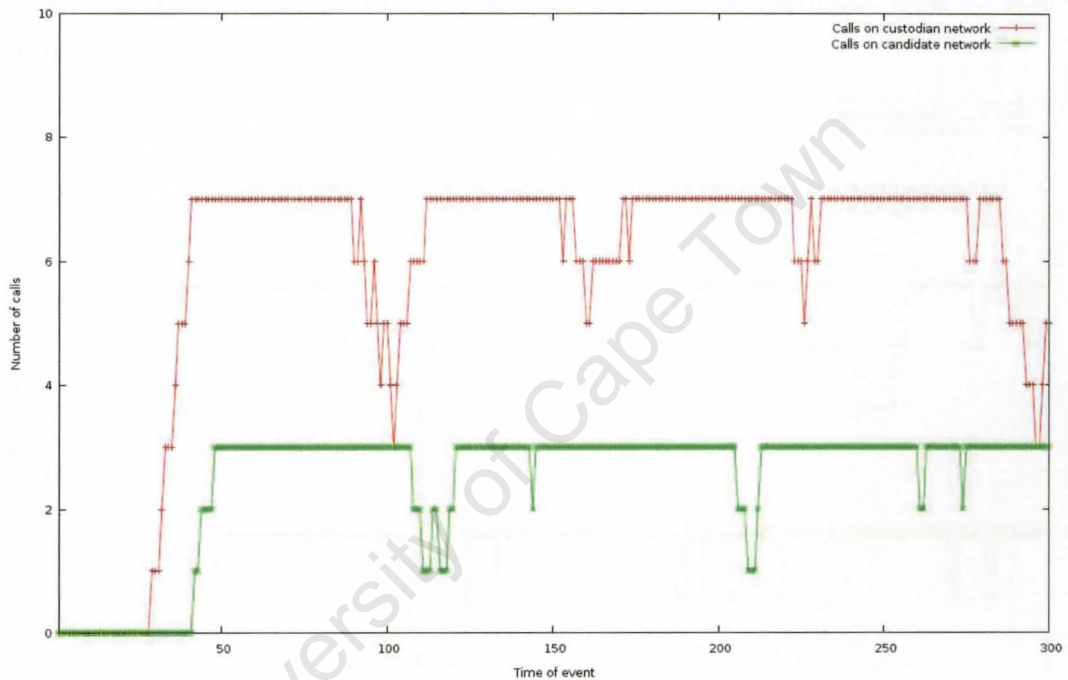


Figure 7.22: The number of custodian and candidate network calls - scenario 3a

The use of a longer call duration leads to an overall better utilisation of the physical capacity on the custodian network. However, during periods of low demand outsourced sessions remain active for a longer period resulting in lost potential income for the custodian network. Applying strategic measures like using shorter outsource durations may reduce the impact of the financial implications on the custodian network.

7.4.6 Scenario 3b

This scenario is characterised by relatively long inter-call and holding periods. The observed trend is illustrated in Fig. 7.23, where outsourcing is triggered during the early phase. When termination of active sessions occurs the physical capacity utilisation falls to low levels, while outsourced sessions remain active. Generally there is a higher degree of fluctuation in the physical capacity as well as the leased capacity utilisation. This results in a high margin of lost potential income due to outsourcing, especially during periods when the occupation of the physical capacity on the custodian network falls to low levels.

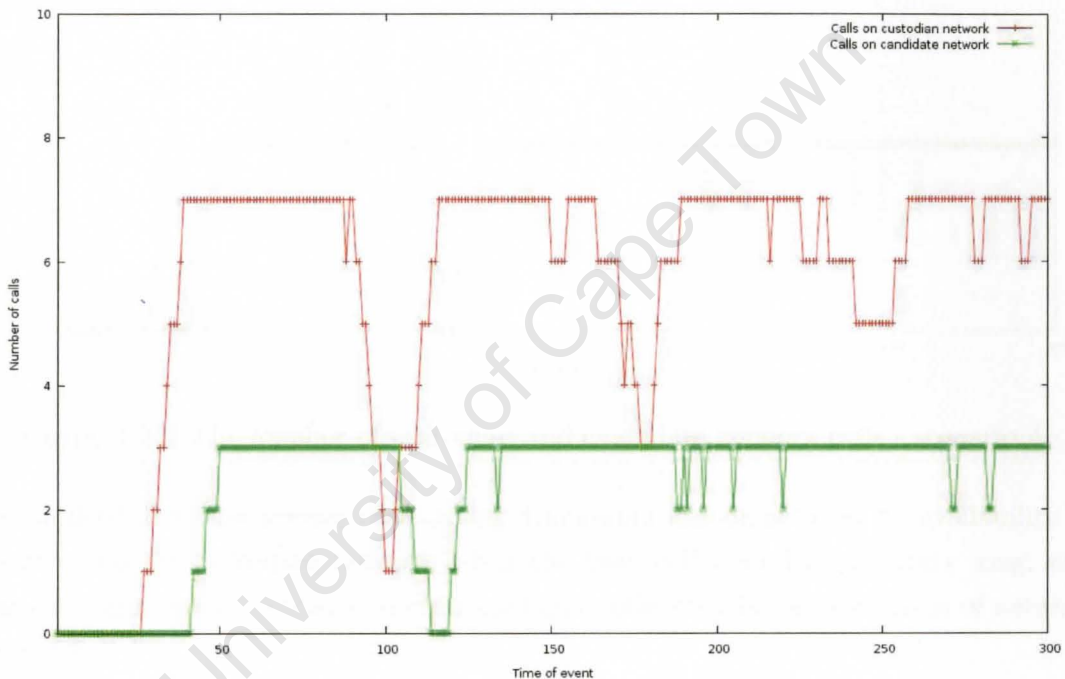


Figure 7.23: The number of custodian and candidate network calls - scenario 3b

7.4.7 Scenario 4

Scenario 4 is characterised by fluctuations in the utilisation of the physical capacity. However, it is observed that the utilisation of the leased capacity responds faster to the availability of unused custodian resources. When compared to scenarios 2a and 2b, there

are more instances when the network loses potential income due to existence of active outsourced sessions while physical capacity is underutilised. The capacity utilisation pattern is shown in Fig. 7.24.

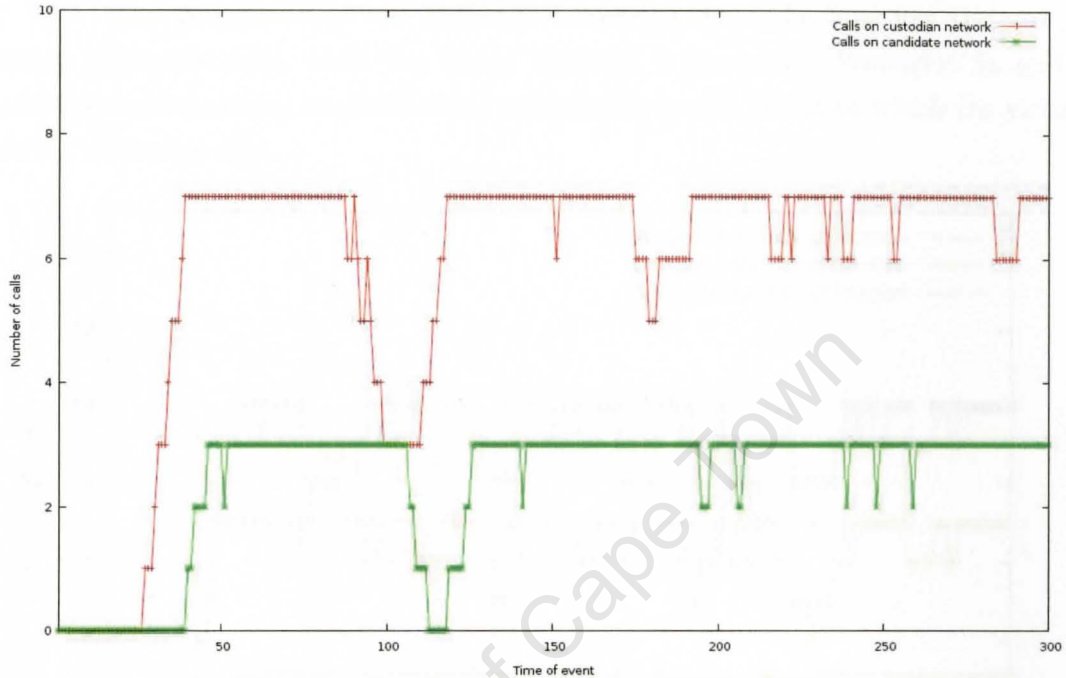


Figure 7.24: The number of custodian and candidate network calls - scenario 4

It is noticed that the service outsourcing framework responds faster to availability of resources on the custodian network when the inter-call period is relatively long, and long call durations are suitable in ensuring higher efficiency in the utilisation of network resources.

7.5 Virtual Capacity Utilisation

The virtual capacity of the custodian network is made of the physical and leased capacity. In the test scenarios discussed above the virtual capacity is 10 sessions. Virtual capacity utilisation refers the total number of sessions being served on custodian and candidate networks at a given time. It highlights the benefits and challenges of service outsourcing when some sessions are served on the candidate network while unused capacity exists on

the custodian network. Scenarios 1a, 3a and 4 are used to illustrate the virtual capacity utilisation.

Scenario 1a is illustrated in Fig. 7.25, scenario 3a in Fig. 7.26 and scenario 4 in in Fig. 7.27. The proportion of lost potential income is shown by how low the virtual capacity utilisation falls, while the leased capacity is occupied. Scenarios 3a and 4 illustrate greater financial implications of outsourcing by the extent to which the virtual capacity utilisation dips.

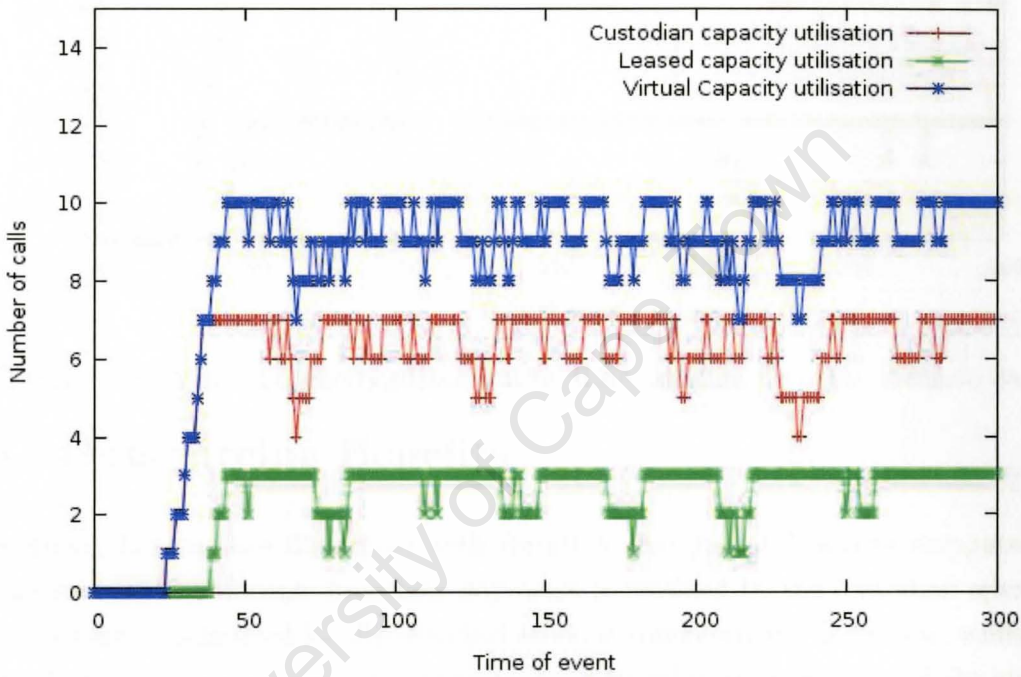


Figure 7.25: Virtual capacity utilisation by the custodian network- scenario 1a

The benefits of service outsourcing revolve around the extra income earned when serving outsourced sessions. No income would be earned if sessions were blocked due to exhaustion of the physical capacity. Results showing the different instances when outsourcing enables the custodian operator to earn extra income are presented.

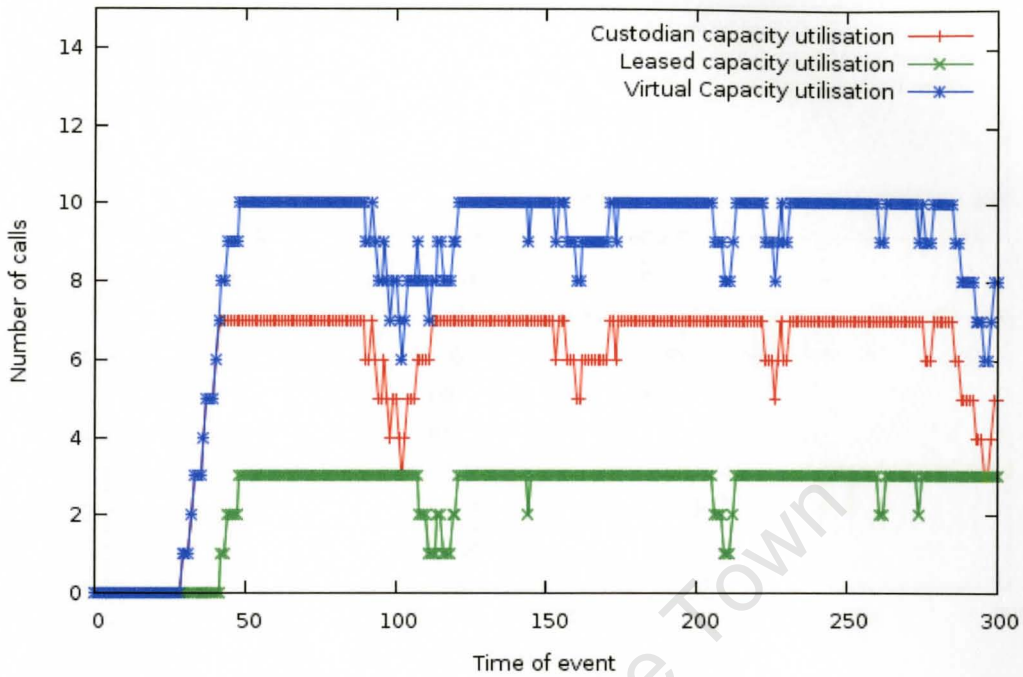


Figure 7.26: Virtual capacity utilisation by the custodian network- scenario 3a

7.6 Outsourcing Benefits

Outsourcing benefits are illustrated with duration charging and income computed at call arrival. 100% of income from user payments is retained by the custodian operator when a session is admitted locally. Blocked sessions contribute 0% of income, while the income from outsourced sessions is inversely proportional to the utilisation of the virtual capacity. This results in lower outsource income when the utilisation is high.

The income retained by the custodian operator for scenario 1a is shown in Fig. 7.28, while the results for scenario 1b are shown in Fig. 7.29.

The graphs show the instances when service outsourcing enables the custodian operator to earn extra income. The points where the unit income per session is zero correspond to the blocking of arriving calls. The income pattern for scenario 2a is shown in Fig. 7.30, while scenario 2b is shown Fig. 7.31.

When compared to scenarios 1a and 1b scenarios 2a and 2b have fewer instances where outsourcing income is earned. However, this extra income contributes to the overall operator income. It is seen that the call duration has an impact on the number of calls

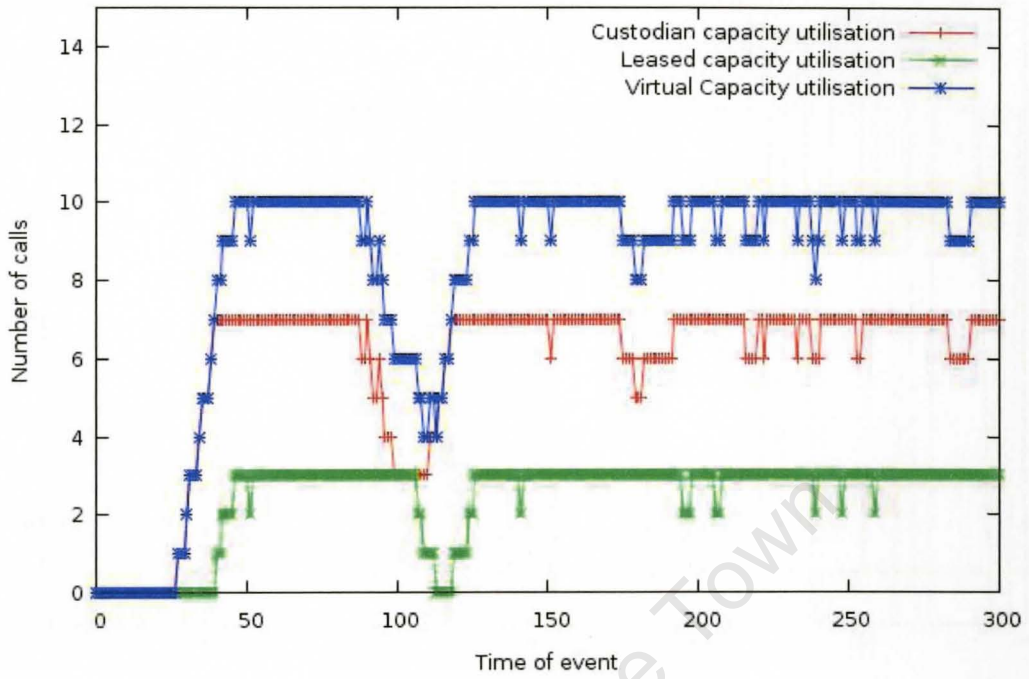


Figure 7.27: Virtual capacity utilisation by the custodian network- scenario 4

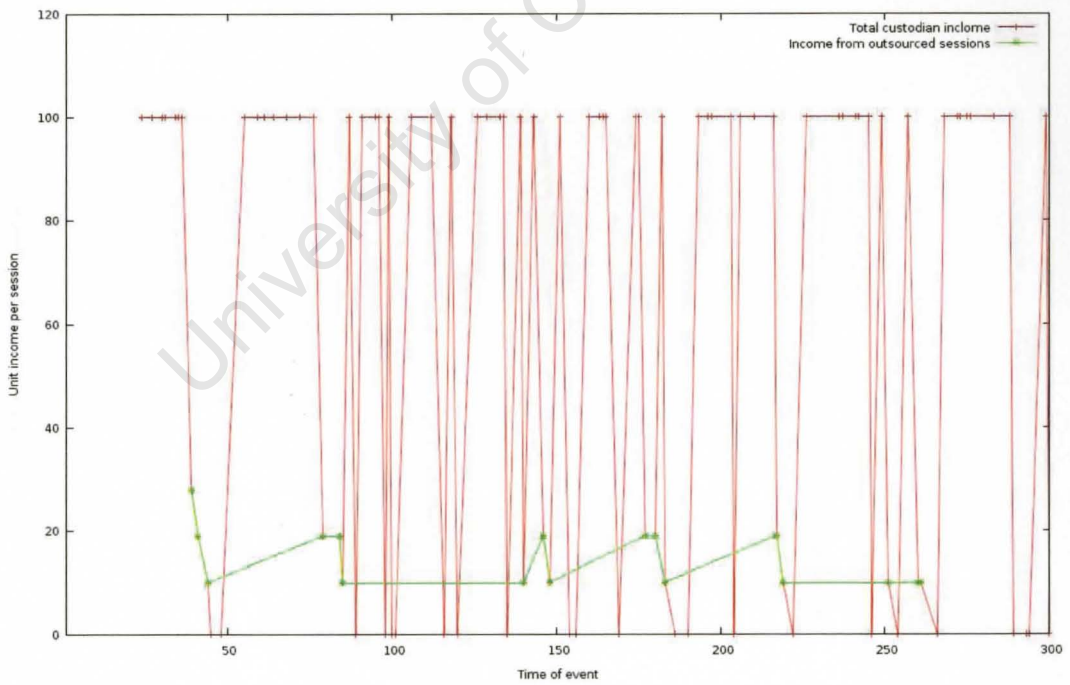


Figure 7.28: Expected income at call arrival - scenario 1a

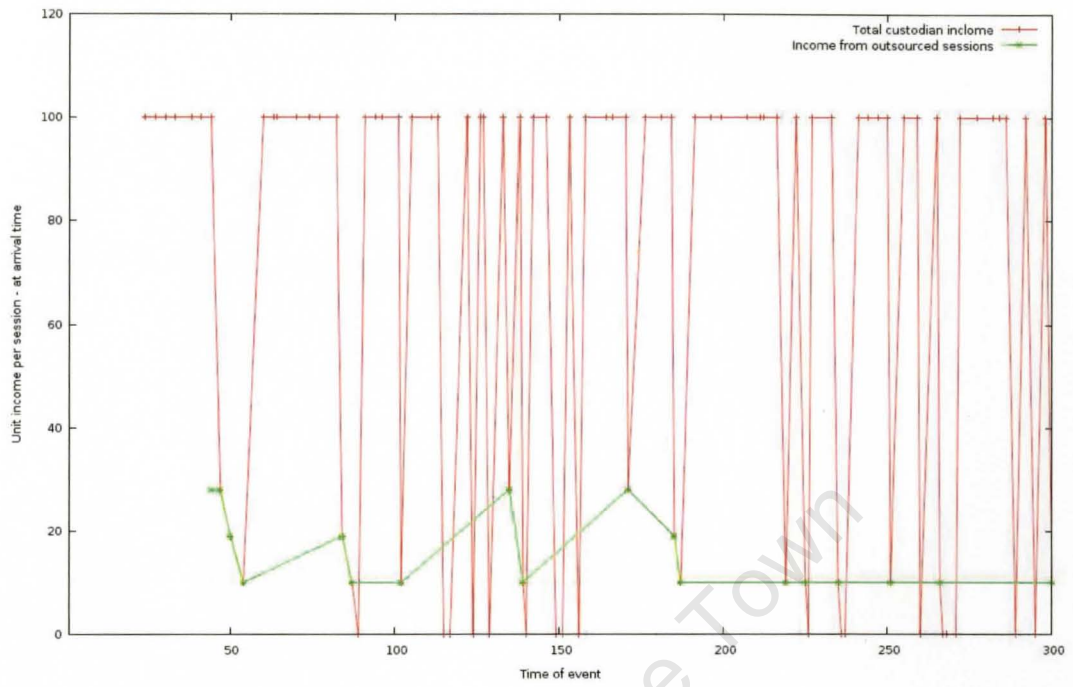


Figure 7.29: Expected income at call arrival - scenario 1b

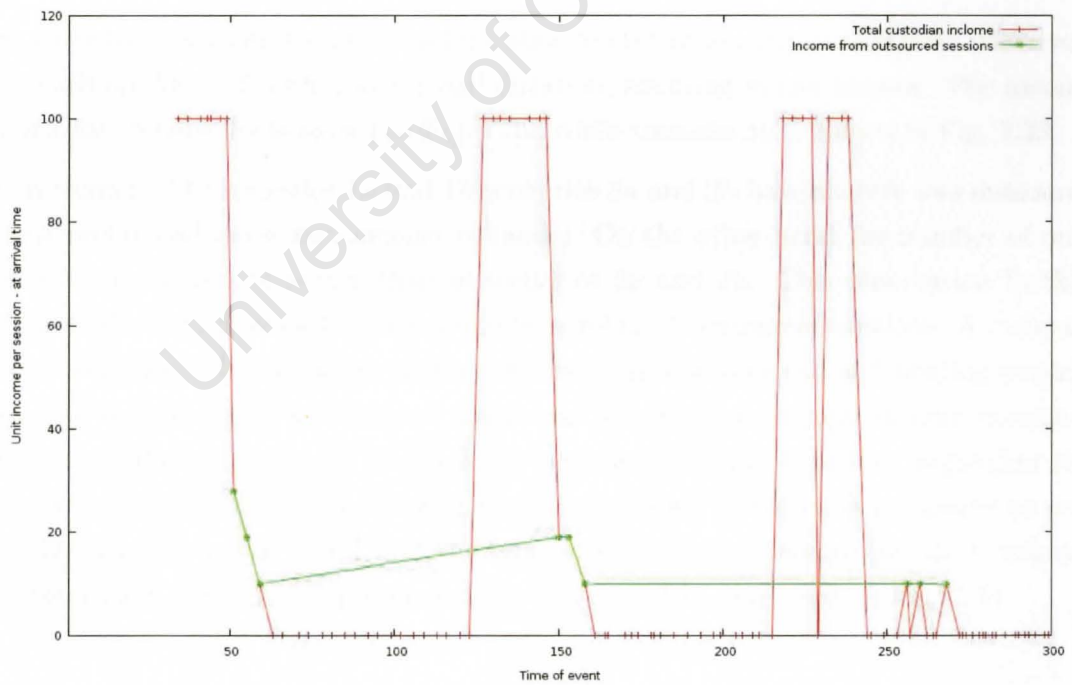


Figure 7.30: Expected income at call arrival - scenario 2a

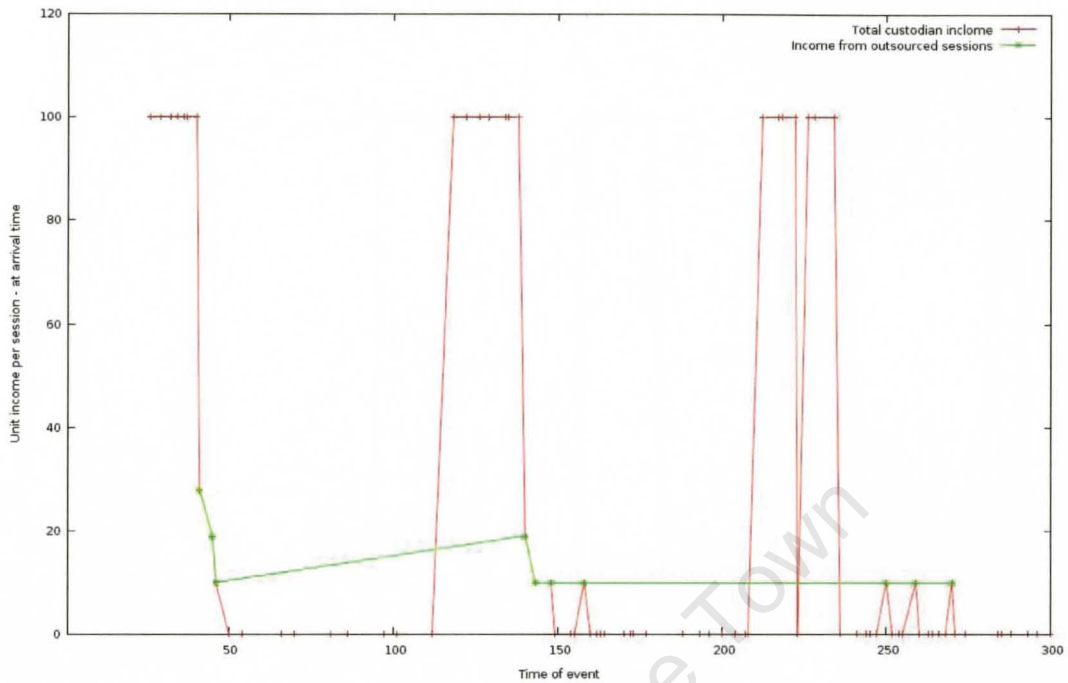


Figure 7.31: Expected income at call arrival - scenario 2b

that are outsourced, since a longer call duration leads to fewer outsourced calls. However, more calls are blocked with a longer call duration, resulting in lost income. The income pattern for scenario 3a is shown in Fig. 7.32, while scenario 3b is shown in Fig. 7.33

When compared to scenarios 1a and 1b scenarios 3a and 3b have slightly less outsource income points and more zero income instances. On the other hand the number of outsource income points is higher than in scenarios 2a and 2b. This observation further highlights the impact of call duration on the number of outsourced sessions. A comparison between scenario 3a and 3b shows that the values of inter-call and holding periods affect the occurrence of zero income instances. There are more zero income instances with lower values of inter-call and holding periods. Scenario 4 further highlights the effect of the holding period by showing more zero income instances as compared to scenario 3b. However, it is noted that the number of outsource income instances reduces with the length of the holding period. This is depicted by the graph in Fig. 7.34.

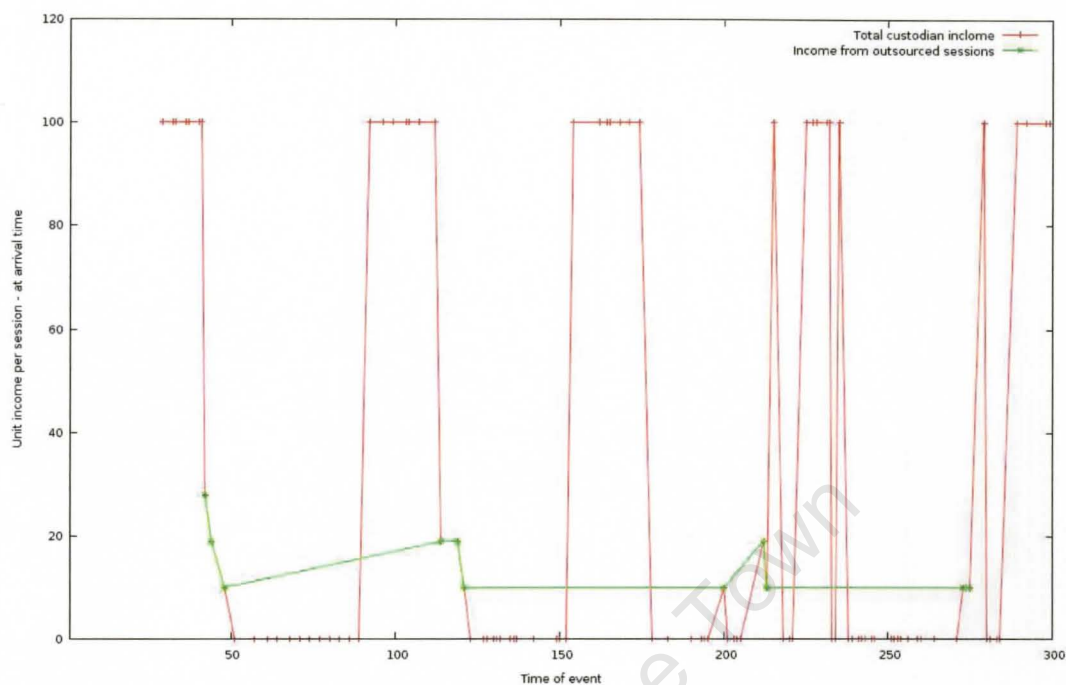


Figure 7.32: Expected income at call arrival - scenario 3a

7.7 Summary

This chapter presented various tests conducted and results of using the outsourcing and charging framework. They highlight the technical impact of outsourcing on IMS procedures. Moreover, evaluation results that demonstrated the performance of the outsourcing framework in terms of the treatment of arriving session establishment attempts were presented. The impact of different call parameters like session duration, inter-call periods and blocked call holding periods on the treatment received by arriving calls was also demonstrated and discussed.

The benefits of service outsourcing to the custodian operator were demonstrated as well as the financial impact of service outsourcing on custodian operator incomes through real-time observation of the utilisation of the virtual capacity. These tests and results are not exhaustive, hence there is room for further research. An analysis of the impact of service outsourcing on candidate network incomes is an area for further investigation.

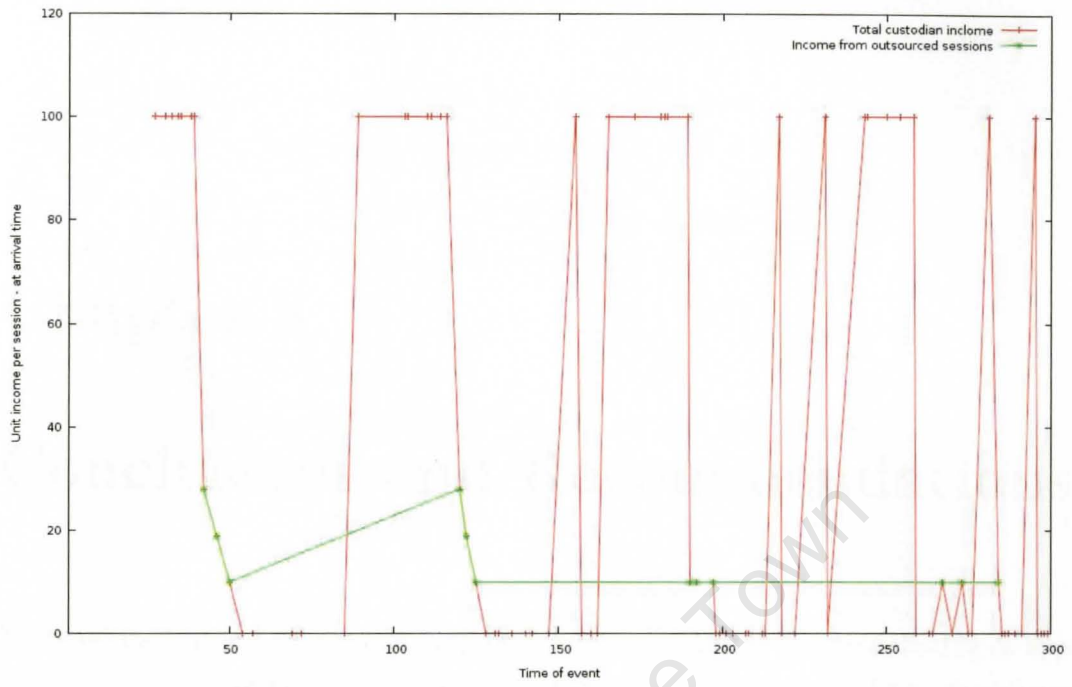


Figure 7.33: Expected income at call arrival - scenario 3b

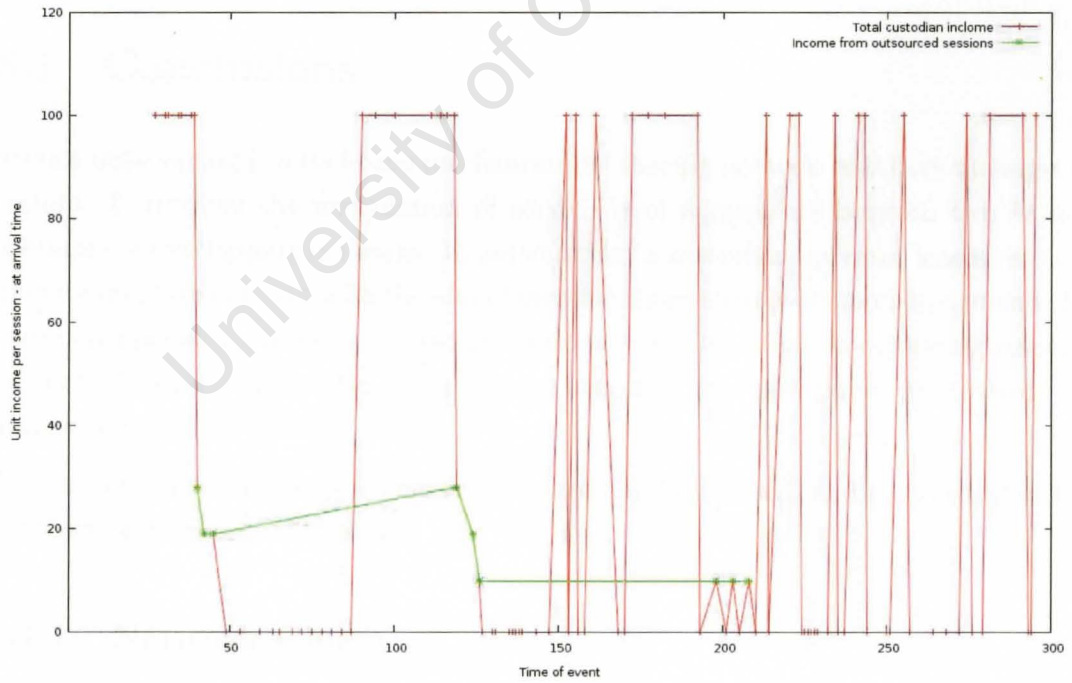


Figure 7.34: Expected income at call arrival - scenario 4

Chapter 8

Conclusions and Recommendations

In the previous chapters an in depth presentation of real-time resource sharing as applied to IMS networks was given. This chapter presents conclusions from the work and a summary of thesis contributions. Areas for further research are identified and recommended.

8.1 Conclusions

Service outsourcing is a real-time mechanism for sharing network resources amongst operators. It involves the formulation of service level agreements between two or more operators of overlapping networks. In outsourcing, a custodian operator leases resources from a candidate network with the aim of meeting short-term overshoots in demand. Operators participating in service outsourcing do so with the intention of sharing resources for mutual benefit. They also share revenue accruing from user payments towards outsourced services.

The concept of service outsourcing is introduced and discussed, and demonstrated with reference to IMS communications.

8.1.1 Network Churn

User experience with network services may determine if they maintain subscriptions with their operator or if they will churn away. Network churn is very expensive for operators, since acquiring new subscriptions is a costly exercise. In the first chapter it is revealed

how outsourcing can aid in ensuring that users have access to network services at any place and anytime, hence contribute to mitigating and minimisation of network churn.

8.1.2 Network Sharing Solutions

In the introductory chapter a discussion of the factors that influence the supply and demand of network resources is given. The supply of network resources is directly related to the costs incurred in providing the necessary infrastructure. On the other hand the demand for network resources is influenced by bandwidth intensive applications. Networks are dimensioned to meet a certain demand level. Upgrading of networks to provide more bandwidth and support higher demand for resources is only feasible in the long term. On the other hand network sharing solutions offer a quicker alternative and can adapt to short-term surges in demand for bandwidth.

In chapter two it is mentioned that network sharing is often used to overcome the lack of network coverage in specific geographical locations, with conventional roaming being an example. Another example is mobile virtual network operators, who provide services to subscribers by utilising infrastructure owned by other operators. These two schemes rely on long term agreements for financial settlement and service cost. These long time frames are not suitable in responding to changes in the demand for resources or congestion levels. Service outsourcing is designed to consider real-time changes in network utilisation in the pricing of resources leased to sharing networks.

8.1.3 Service Outsourcing

Service outsourcing involves real-time formulation of SLAs for resource sharing, it requires a wireless overlap between custodian and candidate networks and guarantees a predictable network agnostic service charges for users. Essentially the capacity of the custodian network is virtually expanded to serve more users. The benefits of service outsourcing were discussed and potential challenges highlighted. The main benefit is that custodian and candidate operators are able to earn extra income from sharing network capacities. The potential challenges revolve around the need to share revenue from user payments. This is illustrated through analytical methods and tests conducted on a testbed.

8.1.4 Operator Attitudes

In the formulation of service outsourcing mechanisms strategies adopted by operators to overcome and minimise associated financial implications are presented. Two broad behavioural categories i.e., optimistic and pessimistic are depicted by operators. Optimistic operators take advantage of all outsourcing opportunities to maximising on the benefits, while pessimistic operators shy away from outsourcing to minimise potential financial implications. The use of shorter outsource periods is proved to be a strategy that pessimistic operators can adopt to minimise outsourcing risks, while benefiting from outsourcing. Moreover, adjustment of the outsource offer price with changing network utilisation level may be used.

8.1.5 Service Outsourcing in IP multimedia Subsystem

The IMS has been standardised as an enabling platform for next generation networks for the delivery of IP based multimedia services. IMS standardisation work is reviewed in chapter 4. The use of service outsourcing within IMS is explored in this thesis using IPTV services and IMS charging procedures. The design and implementation of a service outsourcing and charging framework based on IMS is presented in the fifth and sixth chapters.

8.1.6 Charging in IMS Frameworks

Due to the need for IMS charging frameworks for use in research and standards based charging system that supports online and offline charging is implemented in this work. Details of the design are discussed to reveal how it may be integrated with different IMS services like IPTV and volume based charging systems.

8.1.7 Testbed Implementation

The popularity of Open IMS testbeds in academia and research has risen in recent years. Open Source IMS software from the Fraunhofer institute is used to deploy and test outsourcing by extending the UCT IPTV application server and client. The implementation

comprises an IMS core, an IPTV server based on the UCT advanced video on demand system, the UCT IMS client, a charging system and an outsourcing framework.

Using the framework IPTV sessions are subjected to three kinds of treatment: session establishment attempts are either admitted to the custodian network, outsourced to the candidate network or they are blocked due to exhaustion of virtual capacity required to accommodate them. The implementation provides a proof of the functioning of the outsourcing and charging framework and presents an important point for further research in real-time resource sharing and charging for multimedia services.

8.1.8 Evaluation Tests

The outsourcing and charging framework is subjected to a series of tests to establish the effect of adding new functionality to IMS session setup procedures. Proof of concept tests aimed at tracing the path used by SIP session establishment and registration messages.

Session Establishment Delay

It is proved that charging procedures, both at service and media planes do not contribute significantly to session establishment delay. However, service outsourcing increases the session establishment delay. Sessions that are outsourced to the candidate network incur double the amount of delay when compared to those admitted on the custodian network. The session establishment delay incurred by blocked sessions varies depending on the earliest arising opportunity for admission or outsourcing. This is affected by factors like the duration of active sessions.

Call Acceptance, Outsourcing and Blocking Rates

The test results depict the effect of call duration and call retry intervals on the number and distribution of accepted, outsourced and blocked calls. Generally, longer call durations lead to a high prevalence of blocked calls. This is accompanied by fewer occurrences of accepted and outsourced calls. On the other hand, short call durations lead to a fair distribution of accepted, outsourced and blocked call instances. Moreover, call retry intervals have an effect on the call treatment pattern. Short retry values for blocked calls mean that more attempts are made sooner, hence resulting in a higher

prevalence of blocked calls. This setting is associated with impatient users who would make more attempts to establish blocked calls. These users pose a high risk of churning from networks with high levels of congestion and many blocked calls.

Virtual Capacity Utilisation

Test results reveal the utilisation of the custodian operator's virtual capacity and how it affects the call duration and retry intervals. Although long call duration lead to a steady use of the virtual capacity, the network suffers from larger margins of lost potential income. The loss of potential income is related to the financial implications of outsourcing. On the other hand, short call durations led to a high fluctuation in the utilisation of the virtual capacity. This is further affected by call retry intervals, which influence how fast the system recovers from the income loss phase. In conclusion shorter retry intervals are favourable for this recovery.

8.2 Future work

This thesis has laid down a foundation for investigation into mechanisms for real-time session based resource sharing in mobile networks. There is room for further work in this area; the following aspects have been identified as features of service outsourcing may be investigated further:

8.2.1 Impact of Outsourcing on the Candidate Network

The financial implications and call handling tests performed in this thesis are conducted on the custodian network only, hence there is a need to conduct similar tests to establish the effect of outsourcing on the candidate network. A comparison between results obtained on the candidate and custodian networks will highlight the benefits of outsourcing to each operator. Moreover, further tests are needed to establish the impact of signalling optimisation on service outsourcing transactions.

8.2.2 Practical Deployment

Practical deployment of outsourcing in mobile environments will benefit from enhanced resource monitoring and management frameworks. The IMS outsourcing framework requires a de-registration by the client before handing over to the custodian network. A practical telecommunications environment may have a solution that mitigates the need for de-registration with the custodian network: thus, outsourcing should be tested with practical telecommunication systems.

This will give operators an experience of the benefits of the scheme and encourage them to deploy into their networks.

8.2.3 Application to other Environments

The outsourcing demonstrations presented in this thesis are centred on IMS procedures. There is a need to analyse the usability of outsourcing in emerging network infrastructure deployment schemes, e.g the use of femto cells. Femto cells have been identified as solution to overcoming poor radio signal coverage in indoor environments. Based on the technical and business framework for their deployment, operators could share the resources availed by femto cells using service outsourcing mechanisms.

With respect to testbed implementations there is need to develop and test service outsourcing frameworks on other operating system platforms like windows. Moreover, interoperability of the outsourcing and charging framework with other multimedia service delivery platforms should be investigated as well.

8.2.4 Security and Privacy

Service outsourcing involves the exchange of subscriber information between the custodian and candidate networks. This requires high levels of trust between involved operators, and may expose users to security vulnerabilities like spoofing. Security is an important aspect of telecommunications. The author considers the integration of security in service outsourcing mechanisms an avenue for future research.

Bibliography

- [1] Vitalis G. Ozianyi, Vitor Jesus, Susana Sargento, Rui Aguiar, Neco Ventura, "Virtual Network Capacity Expansion Through Service Outsourcing," *IEEE Wireless Communications and Networking Conference*, pp. 3208–3213, Mar-Apr 2008.
- [2] Claes Beckman, Gregory Smith, "Shared Networks: Making Wireless Communications Affordable," *IEEE Wireless Communications*, pp. 78–85, April 2005.
- [3] Vitalis G. Ozianyi, "A Novel Pricing Approach to Support QoS in 3G Networks," Master's thesis, University of Cape Town, June 2006. MSc Thesis.
- [4] Gonzalo Carmarillo and Miguel Garcia, *The 3G IP Multimedia Subsystem (IMS) - merging the Internet and cellular worlds*. John Wiley and sons Ltd., 2006.
- [5] Miikka Poikselka et. al, *The IMS - IP Multimedia Concepts and Services*. John Wiley, 2 ed., 2006.
- [6] J. Rosenberg et.al, "SIP: Session Initiation Protocol." *RFC 3261*, 2002.
- [7] P. Calhoun et.al., "Diameter Base Protocol." *IETF RFC 3588*, Sept. 2003.
- [8] 3GPP, "Policy and Charging Control (PCC) over S9 reference point," *3GPP TS 29.215 rel. 8*, May 2009.
- [9] P. Weik, D. Vingarzan, T. Magedanz, "The FOKUS Open IMS Core - A Global Reference Implementation," *The IMS Handbook: Concepts, Technologies and Services*, pp. 113–132, Nov. 2008.
- [10] Z. Ezzianer, "Charging and pricing challenges for 3G system," *IEEE Commn. Surveys*, vol. 7, pp. 58–68, Forth Quater 2005.

- [11] 3GPP, "Charging Architecture and Principles - rel.9," *TS 32.240, 3GPP*, Dec. 2009.
- [12] 3GPP, "Policy and Charging Control Architecture - rel.9," *3GPP TS 23.203*, Dec. 2009.
- [13] D. Waiting, R. Good, R. Spiers, N. Ventura, "Open Source Development Tools for IMs Research," *4th International Conference on testbeds and Research Infrastructures for the Development of Networks and Communities (Tridentcom'08)*, Mar. 2008.
- [14] Vitalis G. Ozianyi, Neco Ventura, Eugene Golovins, "A novel pricing approach to support QoS in 3G networks," *Comput. Netw.*, vol. 52, pp. 1433-1450, May (2008).
- [15] Vitalis Ozianyi, neco Ventura, "Real-time Resource Sharing Through Service Outsourcing," *Proceedings of the Local Computer Networks International workshop on the Performance and Management of Wireless and Mobile Networks (P2MNET)*, Denver, USA, October 2010.
- [16] Joyce Mwangama, Vitalis Ozianyi, Neco Ventura, "Charging and billing for composite services in a multi-service provider environment: the IMS case," *Proceedings of the IEEE Wireless Communication Networks Conference (WCNC)*, Sydney, Australia, Apr. 2010.
- [17] Vitalis G. Ozianyi, Neco Ventura, "Design and Implementation of Scalable IMS Charging Systems," *Proceedings of the Local Computer Networks International workshop on the Performance and Management of Wireless and Mobile Networks (P2MNET)*, Zurich, Switzerland, Oct. 2009.
- [18] Vitalis G. Ozianyi, Neco Ventura, "Efficient and scalable charging systems for IP multimedia networks," *IEEE Africon conference, Nairobi, Kenya*, Sep. 2009.
- [19] V. G. Ozianyi, R. Good, P. Wilson and Neco Ventura, "Optimized IP-CANs to support best charged IMS scenario," *Proceedings of Personal, indoor and Mobile Radio Communications Conference, Tokyo, Japan*, September 2009.
- [20] Vitalis Ozianyi, Neco Ventura, "Service outsourcing and billing in inter-domain IMS scenarios," *South African Telecommunications Networks and Applications Conference (SATNAC 2009)*, Mbabane, Swaziland. Sep. 2009.

BIBLIOGRAPHY

- [21] V. Ozianyi, R. Good, N. Carrilho, N. Ventura. "An XML driven framework for policy based QoS management in IMS networks." *Proceedings of IEEE Global communications conference (Globecom)*, New Orleans, LA, USA, Nov - Dec 2008.
- [22] Vitalis G. Ozianyi, Neco Ventura. "Service outsourcing to support virtual home environments (VHE)," *Proceedings of the Local Computer Networks International workshop on the Performance and Management of Wireless and Mobile Networks (P2MNET)*, Montreal, QC, Canada, Oct. 2008.
- [23] S. Blake et. al, "IETF RFC 2475 - An Architecture for Differentiated Services," *IETF RFC*, Dec. 1998.
- [24] N. Blefari-Melazzi, D. Di sorte, G. Reali, "Accounting and pricing: a forecast of the scenario of the next generation Internet." *Elsevier Computer Communications*, vol. 26, pp. 2037-2051, December 2003.
- [25] X. Wang, H. Schulzrinne, "Pricing network resources for adaptive applications in a differentiated services network," *Proc. IEEE INFOCOM*, pp. 943-952. April 2001.
- [26] Bob Briscoe, "A market managed multiservice Internet," *Computer Communications*, vol. 26, pp. 404-414, Mar. 2003.
- [27] J. De Vriendt, P. Laine, C. Lerouge, and X. Xu, "Mobile network evolution: A revolution on the move," *IEEE Communications Magazine*, vol. 40, pp. 104-111, April 2002.
- [28] S. Dixit, Y. Guo, and Z. Antoniou, "Resource management and quality of service in third-generation wireless networks." *IEEE Communications Magazine*, vol. 39, pp. 125-133, February 2001.
- [29] H. Holma, A. Toskala, K. Ranta-aho, and J. Pirkanen, "High-speed packet access evolution in 3GPP release 7," *IEEE Communications Magazine*, vol. 45, pp. 29-35, December 2007.
- [30] H. Eksrom, A. Furuskar, J. Karlsson, M. Meyer, S. Parkvall, J. Torsner, and M. Wahlqvist, "Technical solutions for the 3G long-term evolution," *IEEE Communications Magazine*, vol. 44, pp. 38-45. March 2006.

- [31] D. N. Knisely, T. Yoshizawa, F. Favichia, "Standardization of Femtocells in 3GPP," *IEEE Communications Magazine*, vol. 47, pp. 68–75, September 2009.
- [32] K. Ahmavaara, H. Haverinen, and R. Pinhna, "Interworking architecture between 3GPP and WLAN systems," *IEEE Communications Magazine*, vol. 41, pp. 74–81, November 2003.
- [33] B. Li, Y. Qin, C. P. Low, and C. L. Gwee, "A survey on mobile WiMAX," *IEEE Communications Magazine*, vol. 45, pp. 70–75, December 2007.
- [34] Nokia Siemens Networks, "Network Sharing Solutions." Nokia Siemens Networks - Telkom Access Days, Sept 2009.
- [35] F. Daoud, S. Mohan, "Strategies for provisioning and operating VHE services in multi-access networks," *IEEE Communications Magazine*, vol. 40, pp. 78–88, January 2002.
- [36] L. Bos and S. Leroy, "Toward an all-IP-based UMTS system architecture," *IEEE Network*, vol. 15, pp. 36–45, January/February 2001.
- [37] I. Roussaki, M. Chantzara, S. Xynogalas, M. Anagnostou, "The virtual home environment roaming perspective," *Proc. IEEE ICC*, pp. 774–778, May 2003.
- [38] E. Gustafsson, A. Jonsson, "Always best connected," *IEEE Wireless*, vol. 10, pp. 49–55, Feb. 2003.
- [39] I. F. Akyldiz, S. Mohanty, Jiang Xie. "A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems," *IEEE Comm.*, vol. 43, pp. 29–36, June 2005.
- [40] J. Martin, A. Nilsson, "On service level agreements for IP Networks," *Proc. IEEE INFOCOM*, pp. 855–863, June 2002.
- [41] Panita Pongpaibool, Hyong S. Kim, "Guaranteed Service Level Agreements Across Multiple ISP Networks," *IEEE DRCN*, pp. 325–332, Oct. 2003.
- [42] J. C. Fernandez, T. Taleb, N. Ansari, K. Hashimoto, Y. Nemoto, N. Kato, "Dynamic QoS negotiation for next-generation wireless communication systems," *Proc. IEEE WCNC*, pp. 3058–3062, March 2007.

BIBLIOGRAPHY

- [43] M. Chiang, J. Bell, "Balancing supply and demand of bandwidth in wireless cellular networks: Utility maximization over powers and rates." *Proc. IEEE INFOCOM*, pp. 2800–2811, March 2004.
- [44] C. Thompson, "Supply and Demand Analysis in Convergent Networks," *M.B.A Thesis, MIT*, June 2001.
- [45] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update," *White paper*, pp. 1–8, January 2009.
- [46] D.J. Songhurst, *Charging Communication Networks - From Theory to Practice*. Elsevier, 1 ed., 1999.
- [47] Gonzalo Carmarillo and Miguel Garcia, *The 3G IP Multimedia Subsystem (IMS) - merging the Internet and cellular worlds*. John Wiley and sons Ltd., 2006.
- [48] I. Ali, A. Casati, K. Chowdhury, K. Nishida, E. Parsons, S. Schmid, R. Vaidya, "Network-based mobility management in the evolved 3GPP core network," *IEEE Comm. Mag.*, vol. 47, pp. 58–66, February 2009.
- [49] R. Kuhrie et. al, "Charging in the IP Multimedia Subsystem: A Tutorial," *IEEE Communications Magazine*, vol. 45, pp. 92–99, July 2007.
- [50] J. P. Balbas, S. Rommer, J. Stenfelt, "Policy and charging control in the evolved packet system," *IEEE Comm. Mag.*, vol. 47, pp. 68–74, February 2009.
- [51] 3GPP, "Policy and Charging Control Architecture," *TS 23.203, 3rd Generation Partnership Project (3GPP)*, March 2009.
- [52] 3GPP, "Policy and charging control over Rx reference point," *3GPP TS 29.214 rel.8*, December 2008.
- [53] 3GPP, "3GPP Evolved Packet Core (EPC): Charging Aspects," *3GPP TR 32.820*, vol. Rel.8, Dec. 2007.
- [54] Richard Good, *Linking session based services with transport plane resources in IP Multimedia Subsystems*. PhD thesis, University of Cape Town, 2009.
- [55] 3GPP, "Vocabulary for 3GPP Specifications," *3GPP TS 21.9.5*, vol. Rel. 9, Mar. 2009.

BIBLIOGRAPHY

- [56] M. Koutsopoulou, A. Kaloxylos, A. Alonistioti, L. Merakos, and K. Kawamura, "Charging, accounting and billing management schemes in mobile telecommunication networks and the internet," *IEEE Commun. Surveys*, vol. 7, pp. 50–58, First Quater 2004.
- [57] H. Hakala, "Diameter Credit Control Application," *IETF RFC 4006*, Aug. 2005.
- [58] Miikka Poikselka, *The IMS - IP Multimedia Concepts and Services in the Mobile Domain*. John Wiley and Sons Ltd., 2004.
- [59] Roos M. Hartman and S. Dutnall, "Critical Issues for Roaming in 3G," *IEEE Wireless Communications*, vol. 10, pp. 29–35, Feb 2003.
- [60] E. Mikoczy et al., "IPTV Services over IMS: Architecture and Standardization." *IEEE Comm. Mag.*, vol. 46, pp. 128–135, May 2008.
- [61] 3GPP, "Diameter charging applications," *TS 32.299, 3GPP*, Dec. 2008.
- [62] 3GPP, "3GPP - Network architecture." *TS 23.002 rel. 9*, December 2009.
- [63] I. Ali, A. Casati, K. Chowdhury, K. Nishida, E. Parsons, S. Schmid, R. Vaidya, "Network-based mobility management in the evolved 3gpp core network," *IEEE Comm. Mag.*, vol. 47, pp. 58–66, February 2009.
- [64] Miikka Poikselka, Georg Mayer. *The IMS - IP mUltimedia concepts and services*. John Wiley, 2009.
- [65] Ubuntu Community, "Ubuntu Linux." <http://www.ubuntu.com>, 2007.
- [66] D. Waiting, R. Good, R. Spiers, N. Ventura, "The UCT IMS Client," *Proc. 1st Open IMS Testbeds Workshop, colocated with Tridentcom '09*, 2009.
- [67] Kwang-Cheng Chen, J. Robert B. de Marca, *Mobile WiMax*. John Wiley and Sons, Ltd, 2008.
- [68] K. Bogineni, R. Ludwig, P. Mogensen, V. Nandlal, V. Vucetic, B.K., Yi, Z. Zvonar, "LTE Part I: Core Network," *IEEE Comm. Mag.*, vol. 47, pp. 40–43, February 2009.
- [69] T. Saso, J. Grega, "Long Term Evolution: Towards 4th generation of mobile telephony and beyond," *International conference on Telecommunication in modern Satellite, Cable and Broadcasting Services*, pp. 91–96, 2009.

BIBLIOGRAPHY

- [70] R. Braden, D. Clark and S. Shenker, "Integrated Services in the Internet Architecture: an Overview," *IETF RFC*, vol. RFC: 1633, 1994.
- [71] 3GPP, "Policy and charging control over Gx reference point," *3GPP TS 29.212 rel.8*, December 2008.

University of Cape Town

Appendix A

Wireless Next Generation Networks

A.1 Heterogeneous Next Generation Networks

A number of wireless networks have been standardized to interwork with 3GPP systems. These wireless networks differ in the used frequency spectrum, transmission range, supported data rates, QoS capabilities etc.

A.1.1 WLAN Access

WLANs are standardized in IEEE 802.11 standards. The WLAN MAC mechanism is contention based, thus subscriber stations compete against each other to access the medium [33]. Moreover, only stations that are close to the AP may achieve performance that is good enough to support applications like VoIP and IPTV. IEEE 802.11 WLANs are often deployed using a de facto ISP industry business model, where the core network is an IP backbone provisioned by the ISP. Various servers providing basic IP connectivity services to application specific services can be attached to the IP backbone. Important servers on the IP backbone include a Dynamic Host Configuration Protocol (DHCP) server, a Domain Naming System (DNS) server and an Access gateway.

The DHCP server is required to configure the wireless terminal's IP stack by assigning a relevant IP address, the DNS IP address etc. The DNS performs Internet Fully Qualified Domain Name (FQDN) to IP address translation [32]. A WLAN access point (AP) is required to bridge between the wireless and fixed Ethernet networks. Moreover,

Authentication, Authorization and Accounting (AAA) functions are required in commercial WLAN deployments. A RADIUS server can be used to facilitate AAA functions. A billing system is required in the IP backbone for charging and billing purposes.

In 3GPP-WLAN interworking, access to the WLAN network depends on 3GPP advanced security and subscriber management mechanisms. The mobile subscriber accesses WLAN resources as a roaming user. Visited network selection when roaming outside the coverage of the home operator is standardized in GSM and UMTS specifications [32]. It should be noted that WLANs may support the transmission of multimedia traffic in lightly loaded conditions. Consequently non-QoS based pricing and charging mechanisms are used for these networks.

A.1.2 WiMax Access

The IEEE 802.16 standard for WiMAX has sub-releases for both fixed and mobile WiMAX. It supports high data rates over long distances. Theoretical maximum data rates are approximately 120Mbps, and transmission distances can reach 70 km. WiMAX uses Orthogonal Frequency Division Multiplexing (OFDM) on the physical layer; the connection oriented MAC layer has inherent QoS capabilities, thus making it capable of supporting multimedia applications. Mobile WiMAX (IEEE 802.16e) is a candidate technology for multimedia communications involving moving users. Point-to-multipoint deployments of IEEE 802.16d are suitable for streaming multi-media applications like IPTV. QoS is achieved in the WiMAX physical layer by assigning different code spreading factors or different numbers of codes. Depending on the QoS class used by a Subscriber Station (SS) it can be granted rights to transmit continuously in the uplink direction. Alternatively the base station grants transmission rights upon receiving explicit requests from the SS.

WiMAX QoS Classes

QoS management in WiMAX involves service flows in different states. Service flows are identified by Service Flow IDs (SFID); at the MAC sublayer a Connection ID (CID) is used. In the downlink direction the BS schedules transmissions to each SS. Bandwidth allocation to individual SS for uplink transmissions of admitted service flows is guided

by four service classes [33]. The QoS parameters of interest are latency, jitter and throughput. The WiMax QoS classes are:

- The Unsolicited grant service (UGS) is designed for real-time CBR traffic e.g., VoIP. It provides fixed size transmission opportunities to obviate the need for requests.
- The Real-time polling service (rtPS) is designed for VBR traffic e.g., streaming videos like IPTV. SS are offered periodic request opportunities to specify bandwidth needs.
- Extended rtPS (ErtPS) is specified in IEEE 802.16e to support dynamic bandwidth size allocation.
- Non-real-time polling service (nrtPS) is for delay-tolerant applications with a minimum data rate, e.g., FTP.
- Best effort (BE) service does not specify any service related requirements.

WiMax Functional Architecture

The WiMax network architecture has been defined by the Network Working Group (NWG). NWG specifies the extensive use of IP and IETF standards for interoperability with other networks. IP connectivity is used between all entities; mobile IP is used as the mobility management mechanism [67]. The WiMax architecture can be represented by the reference model in Fig. A.1 [67]. The architecture consists of mobile stations (MS), access service network (ASN) and connectivity service network (CSN). In the WiMax business structure ASNs are implemented by Network Access Providers (NAP) to provide the radio access. A Network Service Provider (NSP) provides IP connectivity and other services, e.g., IMS based services to subscribers based on SLAs with NAPs. A NSP usually controls a CSN.

A.1.3 The 3GPP Evolved UTRAN (E-UTRAN)

3GPP has standardized an evolved packet system (EPS) with two key components; the Service Architecture Evolution (SAE) and Long Term Evolution (LTE). The standards

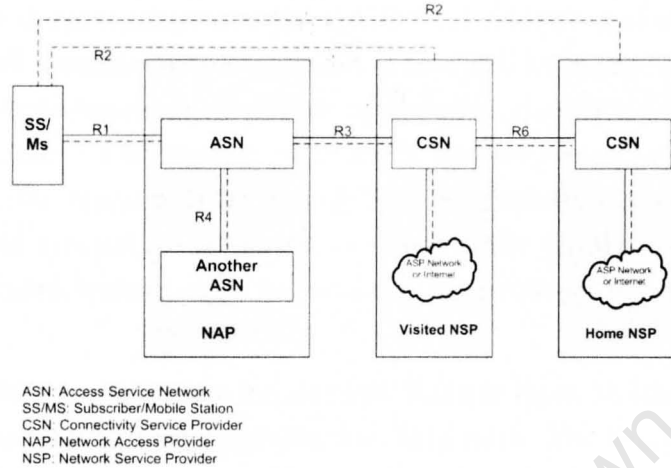


Figure A.1: WiMax network reference Model

provide specifications for an Evolved Packet Core (EPC) and the E-UTRAN. The EPC is a multi-access core network based on IP. It can work with 3GPP (e.g., LTE, UTRAN) and non-3GPP (e.g., WiMax) radio access technologies and fixed line technologies. The EPC provides optimized handovers between different access technologies, and facilitates inter-operator roaming through standardized roaming interfaces. The QoS concept is based on 3GPP's PCC [68].

LTE builds on current 3G networks, and utilizes existing investments. It is designed to co-exist with current 3G access technologies and architectures. LTE creates immense network capacity expansion due to its high data rates. LTE data rates are in the range of 100-300Mb/s and 50-75Mb/s for downlink and uplink transmissions respectively [30, 68]. These data rates come due to improved spectrum efficiency. OFDM access technology is used for the air interface. The flat network architecture contains a reduced number of nodes along the data path. This reduces the number of protocol interfaces, which reduces inter-operability costs. Moreover, LTE reduces overall costs for users and operators.

LTE provides QoS through service differentiation, achieved by packet marking at the UE and Access Core Gateway (ACGW) for uplink and downlink transmissions respectively. Latency reduction supports reliable transport for real-time services, e.g., VoIP. Link layer performance is improved by use of a packet-centric mechanism that combines radio link control and hybrid automatic repeat request protocols [30]. Communication over large cell sizes (upto 120Km radius) is supported through use of few symbols per sub-

frame. Frequency domain adaptation through OFDM achieves performance gain during significant channel variations over bandwidth. The scheduler in the ACGW uses channel quality feedback from terminals to allocate downlink chunks with supported data rates to users. This is achieved by varying power levels, channel coding rates and modulation schemes. Quadrature phase shift keying (QPSK), 16 quadrature amplitude modulation (16 QAM) and 64 quadrature amplitude modulation (64 QAM), modulation schemes are supported. Uplink transmissions use single carrier frequency division multiple access (FDMA).

Multi-antenna solutions, e.g., beam forming and Multiple Input Multiple Output (MIMO) can be used to increase coverage, capacity and data rates. The E-UTRAN connects directly to the EPC. The EPC has the following elements: serving gateway (S-GW), packet data network gateway (PDN-GW) and a mobility management entity (MME) [69]. The EPC supports IP packet switched traffic only. SAE separates the user and control plane; the control plane is managed by the MME. The S-GW performs mobility anchoring in inter-eNodeB handovers. It also supports charging and performs other processing on the packets. Interconnection with other packet networks is achieved via the PDN-GW.

A.2 Differentiated Services Architecture

The DiffServ architecture has been proposed to provide multiple QoS classes in IP networks. Internet applications are divided into elastic and inelastic categories. IMS applications like VoIP and IPTV belong to the inelastic category; thus deterministic QoS guarantees are required to effectively deliver them. IntServ and DiffServ are the main architectures proposed for the provisioning of QoS on the Internet. IntServ relies on per flow Connection Admission Control (CAC) [70]; resource reservation is performed for each IP flow. DiffServ [23] uses aggregated IP flows that are grouped into classes of service. QoS support is facilitated using SLAs between ISPs and their customers.

DiffServ classes of service are priority classes that dictate the treatment of packets at DiffServ routers; traffic in different classes is treated differently. Aggregate flows in a class of service are subjected to the same treatment, known as Per Hop Behaviors (PHB). The Expedited Forwarding (EF) and Assured Forwarding (AF) PHBs have been proposed. The EF PHB requires the departure rate of an aggregate's packets from any

DiffServ node to be equal to or exceed a configurable rate [25]. Four scheduling classes with three drop precedence levels have been defined for the AF PHB.

In general, the performance level of a DiffServ class in terms of packet forwarding depends on the amount of resources allocated to the class, the current load and the congestion level. DiffServ uses SLAs to constrain traffic demand from customers [26]. Traffic conditioning and policing is done to ensure adherence to the terms of the SLA. The performance of DiffServ networks strongly depends on congestion levels. Essentially DiffServ does not have mechanisms for preventing flows in a class of service from overloading the class.

Connection admission control strategies can be used to prevent overloading of traffic classes. To retain the scalability feature of DiffServ admission control should be enforced at the network edge. Flow or connection admission control strategies may be based on first-come-first-served or on the economic potential of flow. In this case flows that can lead to higher income for the operator will be admitted first.

Appendix B

Charging control in IP Multimedia Subsystem

B.1 IMS session control

Every User Agent (UA) registers with the home IMS domain before session establishment. Registration transactions are handled by the P-CSCF, I-CSCF, S-CSCF and HSS. During registration the S-CSCF downloads the user profile from the HSS. The profile contains a list of Public User Identities that can be used for establishing sessions; it also contains other subscription information.

IMS sessions can be established between IMS users or between users and application servers. In user-to-user sessions the caller sends an IMS INVITE towards the callee. The INVITE message traverses a set of CSCFs in the caller and callee domains. The S-CSCF performs an evaluation of initial filter criteria (iFC) to trigger services subscribed to by the user. The iFC is downloaded from the HSS as part of the user profile during registration. Preconditions are used during session establishment to meet requirements of some sessions, e.g. reservation of radio bearer resources for media plane traffic. Figure B.1 gives a simplified illustration of IMS signalling during session establishment. The session participants are assumed to be in the same domain; only a few messages are shown. To establish a session UA1 sends an INVITE towards UA2. UA2 answers with a 200ok to accept the INVITE. To end the session UA1 sends a BYE message to UA2. Routing of SIP session establishment and control messages utilises *Route* and *Record-Route* headers

inserted by the UAs and CSCFs.

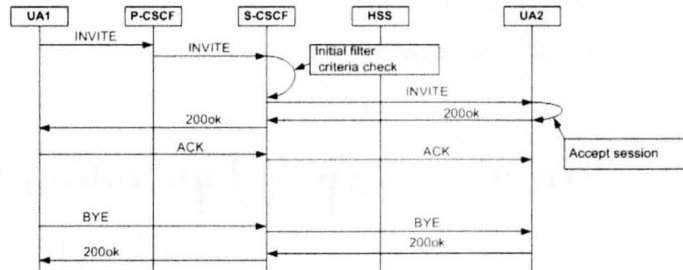


Figure B.1: IMS session control

When application servers are involved, session signalling is intercepted by the S-CSCF for filter criterion evaluation to determine to which AS the messages should be sent. The target AS receives and processes SIP messages and responds to client requests, or performs network specific functions like charging.

SIP and Session description protocol

As mentioned earlier SIP is the main protocol for session management in the IMS. It is a robust text based application layer protocol used to convey different types of information between user agents and servers. The user agent comprises a user agent client, which initiates SIP requests and a user agent server, which receives and processes SIP requests [5]. Several types of SIP servers exist, i.e., proxy, redirect, location and registrar server.

SIP messages contain a header and body; the header contains the request or status line, Via header, initiator and receiver of the request, call ID etc. The request line contains the SIP method (e.g., INVITE and BYE), request URI and SIP version, i.e. 2.0 in the current release. The Status line is used for responses: it contains the SIP version, status code and reason phrase. Status codes have particular meanings, i.e., 1xx are provisional responses, 2xx indicate success, 3xx are redirection responses, 4xx indicate client errors, 5xx indicate server errors and 6xx indicate global failure [5].

The message body can convey any type of text based information. sdp messages are typically conveyed in the SIP body. sdp is a text based application layer protocol. It is used by user agents for session and timing description, and to indicate media formats and RTP address and port numbers. Among the important lines in an sdp body are

'm=', 'b=' and 'a=' lines. The 'm=' lines carry information about the media type, e.g., audio or video and the codecs used; the 'b=' lines convey the bandwidth information; the 'a=' lines are optional lines that provide more information about the media stream [5].

B.2 IMS policy and charging control

B.2.1 Online charging

The Diameter credit control application is used in online charging [57]. Credit control is required to pre-authorise sessions prior to commencement of resource consumption. Similar to offline charging, online charging requires a CTF integrated in an AS. In addition an Online Charging Function (OCF) is required; the OCF is an integral part of the OCS. For online charging the CTF may be integrated in the following IMS entities: MRCF, SIP AS and S-CSCF. An IMS gateway function is required to support credit control at the S-CSCF. Communication between the CTF and the OCF occurs across an Ro reference point. The Ro interface uses the Diameter credit control application. When a chargeable event is detected a CCR message is sent to the OCS. The tariff function in the OCS will determine the value of the requested resources to enable allocation of credits for a specific amount of resources, e.g., duration of usage etc. The allocated credits will be sent to the CTF in a CCA. In the course of resource consumption the account balance management function performs supervision of credit usage in order to signal termination of service delivery if credits are exhausted [11].

B.2.2 Offline charging

The entities involved in offline charging include: a CTF that is located in the AS, a Charging Data Function (CDF), a Charging Gateway Function (CGF) and a billing domain [49, 11]. The CDF and CGF can be integrated into one entity in the billing domain. The Rf reference point inter-connects the CTF and CDF. The CTF generates and sends Diameter Accounting Requests (ACR) to the CDF upon detection of chargeable events. For offline charging, a CTF may be integrated in the following IMS entities: I-CSCF, S-CSCF, P-CSCF, SIP AS etc. The ACR is sent in Diameter AVPs. When the start of a charging session is detected the CTF starts collecting charging events and transfers

the collected information in *interim* ACR messages to the CDF. At the termination of a charging session a final ACR is sent to the CDF. Accounting Answer (ACA) messages are sent by the CDF to the CTF in response to received ACR messages. Offline and online charging functions and interfaces are depicted in Fig. B.2.

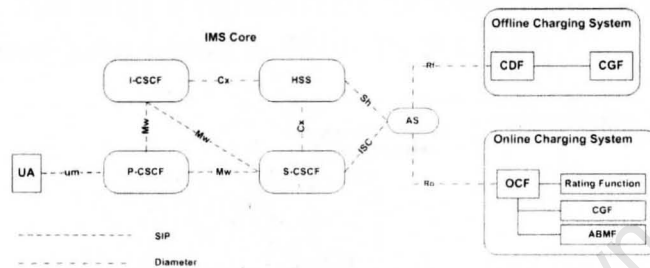


Figure B.2: Service level charging in IMS

B.2.3 IMS Session Based Charging

Session based charging is used when the expected resource usage cannot be determined upfront. It involves the generation of multiple charging events that are sent to the CDF or OCS [49]. The start of a user session is detected at the application server as a chargeable event, which is mapped to a charging event. Unit reservation of credits is required if online charging is used. The reserved credits correspond to the authorised number of minutes or bytes. Credit supervision is performed during the lifetime of a session to ensure timely termination of service delivery if credit exhaustion occurs. In the course of a session interim charging events are issued by the AS to request more credits from the OCS, and a final charging event is issued at session termination. Unused credits will be credited to the user's account.

B.2.4 Policy and charging control over Gx reference point

The GX reference point is located between the PCRF and PCEF [71]. It is used for provisioning rules from the PCRF to the PCEF; removal of provisioned rules occurs across this reference point as well. The Gx reference point is based on Diameter credit control application. Generally PCC rules are installed, modified or removed; pre-defined rules are either activated or de-activated. PCC rules can be requested (pulled) by the

PCEF using a CC-Request; the PCRF responds with a CC-Answer. Unsolicited (push) provisioning is performed by the PCRF by sending an RA-Request message to the PCEF. The PCRF uses the Charging-Rule-Name to identify provisioned rules when PCC rule modification or removal is performed. Figure B.3 depicts PCC functional entities and reference points. This figure is repeated here for reader convenience. Standardisation work on Gx reference point is given in 3GPP TS 29.212 [71].

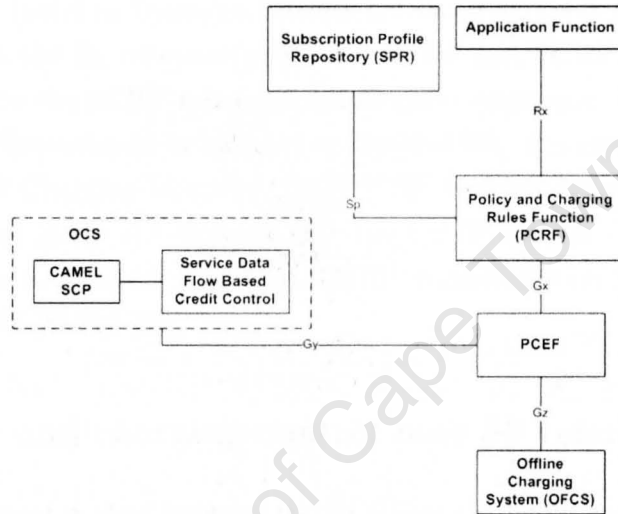


Figure B.3: Policy and charging control architecture

Policy and Charging Rules Function (PCRF)

The PCRF encompasses policy control and flow based charging. It performs network control for data flow detection, gating control and flow based charging [71]. It makes PCC rule decisions based on information obtained from an application function over the Rx reference point, the PCEF over the Gx interface, the SPR over the Sp interface or pre-configured policies in the PCRF.

Policy and Charging Enforcement Function (PCEF)

The PCEF performs policy enforcement and flow based charging. The PCEF is located at the gateway, e.g., GGSN for GPRS networks and PDN GW for EPS. It controls user plane traffic QoS, flow detection and counting [71]. It also performs online and

offline charging interactions for media plane traffic. The PCEF enforces gating control according to received PCC rules. It performs charging control based on PCC rules and credit authorisation by the OCS for online charging.

B.2.5 Policy and charging control over Rx reference point

The Rx interface is based on Diameter network access server application. The functional elements used with the Rx reference point are the AF and PCRF. Session information is sent by the AF to the PCRF using the AA-Request command, with an indication of whether media IP flows should be enabled or disabled [52]. For charging correlation the AF includes an AF-Charging-Identifier. The PCRF performs session binding based on information received in the AA-Request, i.e., the UE IP address. The PCRF responds to the AF using the AA-Answer command. 3GPP standardisation on Rx reference point is given in 3GPP TS 29.214 [52].

B.2.6 Policy and charging control over S9 reference point

The S9 reference point is used between the PCRF in the home domain (H-PCRF) and the PCRF in the visited domain (V-PCRF) [8]. It supports roaming scenarios involving a home and visited mobile network. Both S9 and Rx Diameter applications are used over S9 reference point. S9 Diameter application is used to install PCC and QoS rules generated in the home network to the visited network, and to send events in the opposite direction. If the AF is located in the visited network Rx over S9 is used to send session information from the V-PCRF to the H-PCRF. Figure B.4 illustrates the S9 reference point architectural setting. This is a home routed scenario, where all media plane traffic is sent to the home network. The H-PCRF creates PCC rules based on session information obtained from the AF across Rx interfaces; information obtained from V-PCRF, SPR, PCEF etc is also used.

The V-PCRF receives event triggers from the H-PCRF, for home routed access, and validates them according to operator policies before provisioning them to the BBERF. For visited access it requests PCC rules from H-PCRF and send event reports to H-PCRF. 3GPP standardisation on S9 reference point is available in 3GPP TS 29.215 [8].

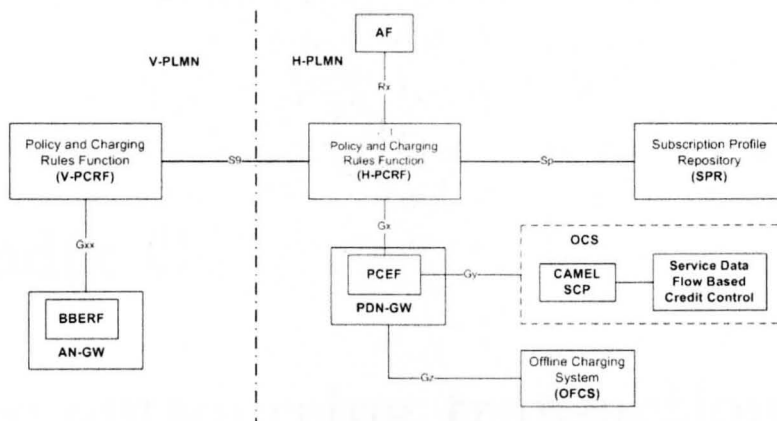


Figure B.4: S9 policy and charging control reference architecture

B.2.7 Real-time Transport Protocol (RTP)

RTP is used for end to end delivery of real-time media. It facilitates codec identification, sequence numbering, timestamping etc. RTP relies on the RTP Control Protocol (RTCP) to provide QoS monitoring and to convey session participant information [5]. In IMS RTP and RTCP are used on the media plane for delivery of media traffic, e.g., VoIP and IPTV.

Appendix C

Service outsourcing transactions

C.1 Numerical data

Table C.1: Numerical data

Duration of sessions 5 minute frame

Network 1

Full network capacity (Number of maximum QoS level sessions) = 20

Session Interval	No. of outsourced sessions	Requested QoS level	Offered per unit price	Negotiated per unit price	Outsource duration	Network load	Potential home network load	Home capacity unused by outsourced sessions	Normal service cost	Income from home sessions	Expected income from home sessions	Income from outsourced sessions	Net Income
0	0	N/A	N/A	N/A	N/A	20	20	0	2	40	40	0	40
1	2	1	1.5	1.6	2	20	20	0	2	40	40	0.8	40.8
2	3	1	1.5	1.7	1	18	20	2	2	36	40	0.9	36.9
3	1	1	1.5	1.5	1	20	20	0	2	40	40	0.5	40.5
4	5	1	1.5	1.5	3	18	20	2	2	36	40	2.5	38.5
5	3	1	1.5	1.6	1	19	20	1	2	38	40	1.2	39.2
6	1	1	1.5	1.8	1	18	19	1	2	36	40	0.2	36.2
7	0	N/A	N/A	N/A	N/A	19	19	0	2	38	38	0	38
8	2	1	1.5	1.6	1	20	20	0	2	40	40	0.8	40.8
9	1	1	1.5	1.5	2	20	20	0	2	40	40	0.5	40.5
10	3	1	1.5	1.5	2	20	20	0	2	40	40	1.5	41.5
11	6	1	1.5	1.6	3	20	20	0	2	40	40	2.4	42.4

Network 2

Full network capacity (Number of maximum QoS level sessions) = 30

Session Interval	No. of outsourced sessions	Requested QoS level	Offered per unit price	Negotiated per unit price	Outsource duration	Home Session requests	Serviced home sessions	Blocked home sessions	Normal service cost	Income from home sessions	Expected income from home sessions	Income from outsourced sessions	Net Income
0	0	N/A	N/A	N/A	N/A	20	20	0	2	40	40	0	40
1	2	1	1.5	1.6	2	25	25	0	2	50	50	3.2	53.2
2	3	1	1.5	1.7	1	23	23	0	2	46	46	5.1	51.1
3	1	1	1.5	1.5	1	24	24	0	2	48	48	1.5	49.5
4	5	1	1.5	1.5	3	22	22	0	2	44	44	7.5	51.5
5	7	1	1.5	1.6	1	25	23	2	2	46	50	4.8	50.8
6	1	1	1.5	1.8	1	28	28	0	2	56	56	1.8	57.8
7	0	N/A	N/A	N/A	N/A	26	26	0	2	52	52	0	52
8	2	1	1.5	1.6	1	20	20	0	2	40	40	3.2	43.2
9	1	1	1.5	1.5	2	20	20	0	2	40	40	1.5	41.5
10	3	1	1.5	1.5	2	22	22	0	2	44	44	4.5	48.5
11	6	1	1.5	1.6	3	30	24	6	2	48	60	9.6	57.6

Appendix D

Outsourcing procedures in IMS

D.1 IMS Registration

IMS registration creates a binding between the SIP address and the IP address of the user. Figure D.1 [47] gives a full illustration of IMS registration transactions.

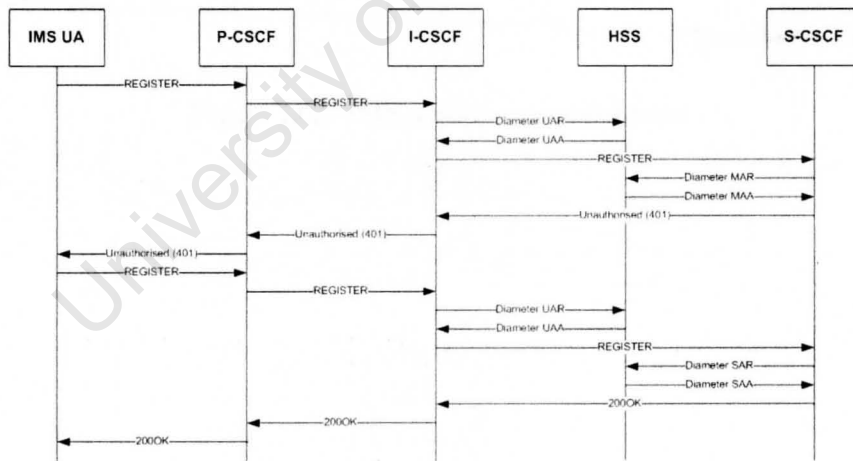


Figure D.1: Registration in the IMS

D.2 Session establishment

A detailed illustration for IMS session establishment between two users, Alice and Bob, belonging to different IMS domains is given in Fig. D.2. This figure is adopted from [64].

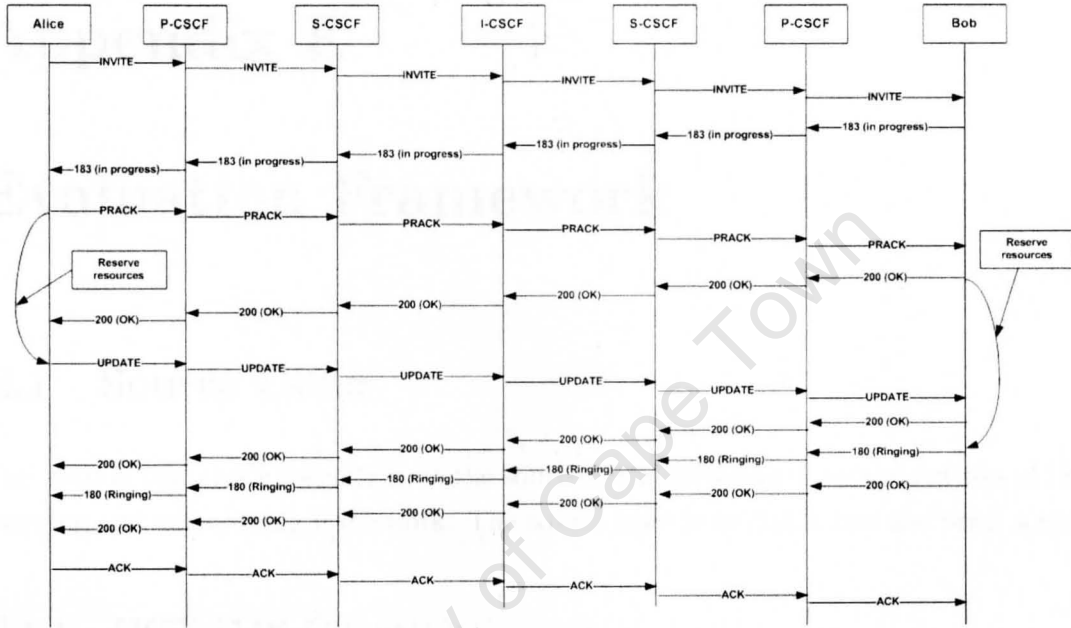


Figure D.2: IMS sessions establishment

Appendix E

Evaluation Framework

E.1 Source Code

This section contains information on the sources of various functions and entities of IMS charging and outsourcing framework. The source code is available free and open source.

E.1.1 UCT IMS Client

The client sources are included in the accompanying CD ROM in the `imsclient` directory. The sources are structured according to the functional architecture of the client.

E.1.2 IPTV Application Server

The source code is included in the accompanying CD ROM in the `iptv-as` folder. The server is started in `main.c` where the `keys_file` containing the TV channels is loaded; the Diameter peer information that identifies the CTF, CDF and OCS is loaded as well. The timers, CTF Diameter process and Diameter listener are started. More core functions are contained in `charging.c`; it contains the timers, call object handlers and sdp manipulation functions. IPTV functions are contained in the `uct_aiptv_server.c` file in the `iptv` folder. The SIP events listener process is contained here.

TV channels

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- This is an example of the file that the server looks
for when starting the UCT IMS IPTV Application Server
The key is any text between the SIP: and the @domain that
the server receives in an invite.
So for example sip:channel1@uct.ims will be parsed to channel1.
The value is the rtsp address of the server along with the path
of the media file the client is requesting.
-->
<key-value_pairs>
<key-value_pair>
<key>channel1</key>
<value>rtsp://tv.imscore.ims:5554/channel1</value>
</key-value_pair>
<key-value_pair>
<key>channel2</key>
<value>rtsp://tv.imscore.ims:5554/channel2</value>
</key-value_pair>
<key-value_pair>
<key>channel3</key>
<value>rtsp://tv.imscore.ims:5554/channel3</value>
</key-value_pair>
</key-value_pairs>
```

Diameter peer configuration

```
<?xml version="1.0" encoding="UTF-8"?>
<DiameterPeer
FQDN="iptv.imscore.ims"
Realm="imscore.ims"
Vendor_Id="10415"
Product_Name="CDiameterPeer"
AcceptUnknownPeers="1"
DropUnknownOnDisconnect="1"
Tc="30"
Workers="4"
QueueLength="8"
>
<Peer FQDN="cdf.imscore.ims" Realm="imscore.ims" port="3892" />
<Peer FQDN="ocf.imscore.ims" Realm="imscore.ims" port="3891" />
<!--<Peer FQDN="hss.imscore.ims" Realm="imscore.ims" port="3868" />-->
<Acceptor port="3871" bind="127.0.0.1" />
<Auth id="16777218" vendor="10415"/><!-- Rf /-->
<Auth id="16777256" vendor="10415"/><!-- Ro unassigned by 3gpp at the moment /-->
<Auth id="16777217" vendor="10415"/><!-- 3GPP Sh -->
</DiameterPeer>
```

HSS configuration

The HSS is configured with iFC and SPT for various services to trigger the redirection of SIP messages to the IPTV AS. Figure E.1 shows the configuration of the iFC, which is mapped to the defined IPTV AS and SPT.

Initial Filter Criteria -iFC-

ID	
Name*	iptv_filter
Trigger Point	IPTV_trigger
Application Server*	iptv
Profile Part Indicator	Any

Mandatory fields were marked with "*"

Figure E.1: IPTV iFC configuration at the HSS

The configuration of the service profile is depicted in Fig. E.2. The iFC is attached to the service profile.

Service Profile -SP-

ID	
Name*	default_sp
Core Network Service Auth	0

Mandatory fields were marked with "*"

Attach IFC

Select IFC... - Priority 0

List of attached IFCs

ID	iFC Name	Priority	Detach
1	default_ifc	0	<input type="button" value="Detach"/>
2	iptv_filter	2	<input type="button" value="Detach"/>

Attach Shared-IFC-Set

Select Shared-IFC...

List of attached Shared-IFC-Sets

ID-Set	Name	Detach
1	default_shared_set	<input type="button" value="Detach"/>

Figure E.2: The HSS service profile configuration

E.1.3 RTSP Media Server

The media server is started by the following command:

```
gnome-terminal -tab -t VLC-Media-Server -e "vlc --ttl 12 -vvv --color -I telnet --rtsp-host tv.imscore.ims:5554 --vlm-conf /home/vitozy/Research/OPEN-IMS.TEST/media-server/Configuration/vlc_vod_config"
```

The configuration file (vlc_vod_config) is included in the accompanying CD ROM in testbed/media-server.

E.1.4 Outsourcing User Agent

The service outsourcing user agent is integrated into the IMS client. The source code of the IMS client is included on the accompanying CD ROM on the `uctimsclient` folder.

E.1.5 Network Agent

The source code of the outsourcing network agent is contained in the outsourcing subfolder of the IPTV folder on the accompanying CD ROM.

E.1.6 Charging Framework

The charging application server uses a SIP module, a Diameter module and functions that interface with the OCS, OFCS and the PCRF. A sample Diameter configuration file is given below. The configuration contains the domain name of each Diameter peer and the port number used for binding.

```
<?xml version="1.0" encoding="UTF-8"?>
<DiameterPeer
  FQDN="charging.imscore.ims"
  Realm="imscore.ims"
  Vendor_Id="10415"
  Product_Name="CDiameterPeer"
  AcceptUnknownPeers="1"
  DropUnknownOnDisconnect="1"
  Tc="30"
  Workers="4"
  QueueLength="16"
  >
  <Peer FQDN="pcef.imscore.ims" Realm="imscore.ims" port="6868"/>
  <Acceptor port="3879" bind="127.0.0.1" />
  <Auth id="16777236" vendor="10415"/> <!-- Rx /-->
  <Auth id="16777236" vendor="0"/>
</DiameterPeer>
```

The source code of the charging framework is included in the `ctf_rf` and `ctf_ro` folders in the IPTV folder on the accompanying CD ROM.

Online Charging System

The source code of the OCS is included in the accompanying CD ROM. It is located in the UCTOCS folder.

Offline Charging System

The source code for the offline charging system is included in the accompanying CD ROM. The files are contained in the UCTCDF folder.

University of Cape Town

Appendix F

Accompanying CD-ROM

The software developed in the course of executing this thesis is available as Free Open Source software. It has been included with this thesis on the accompanying CD-ROM. All files used in preparing the thesis document have also been included on the CD-ROM.

The CD-ROM is structured as follows: A pdf version of the thesis document is included; there is a *Lyx-files* directory containing all lyx and other source files (e.g., figures) used to generate the document; the source code of the testbed software is included in the *Testbed* directory; and a collection of all test results are available in the *Results* directory.